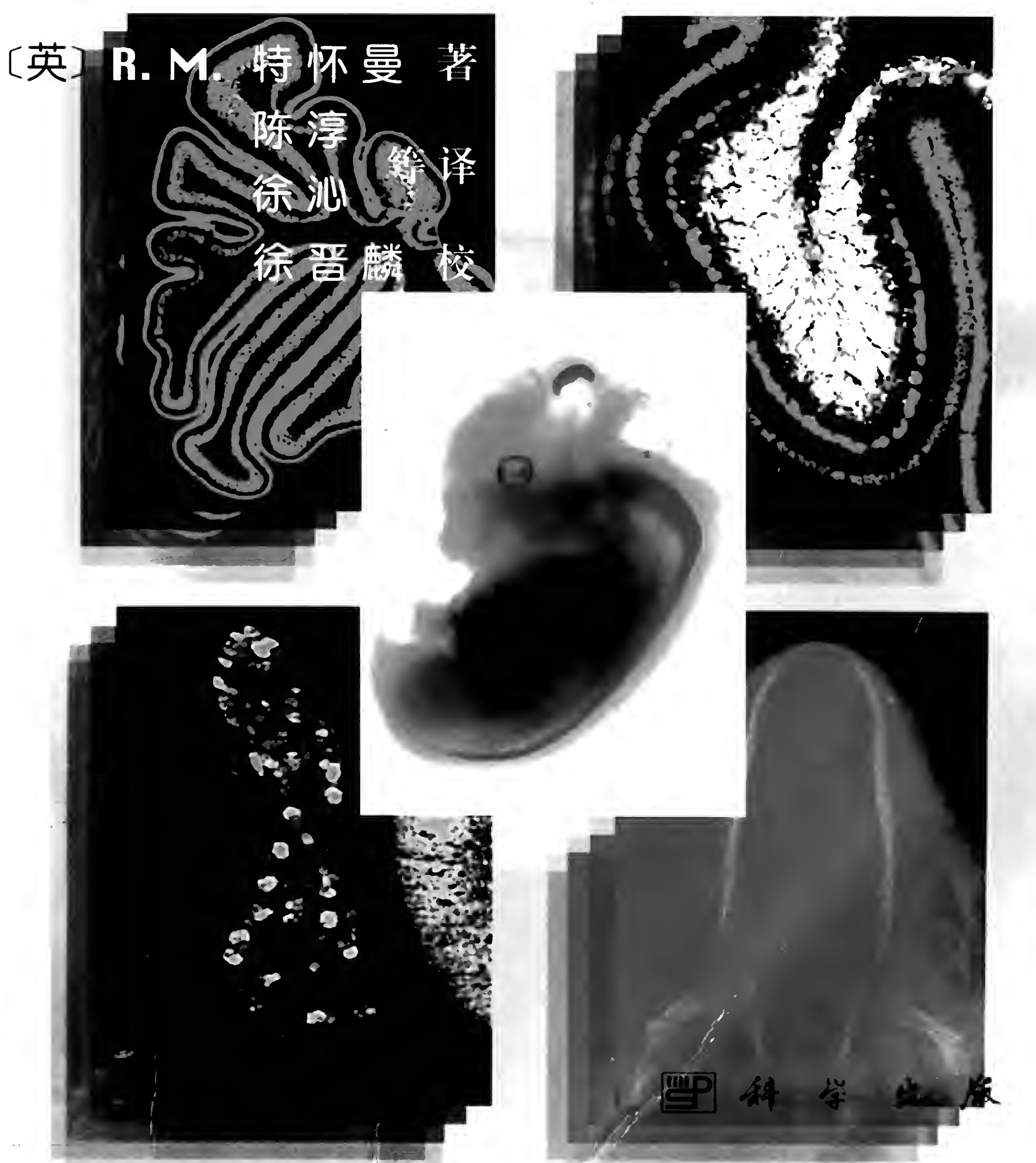
国外优秀教

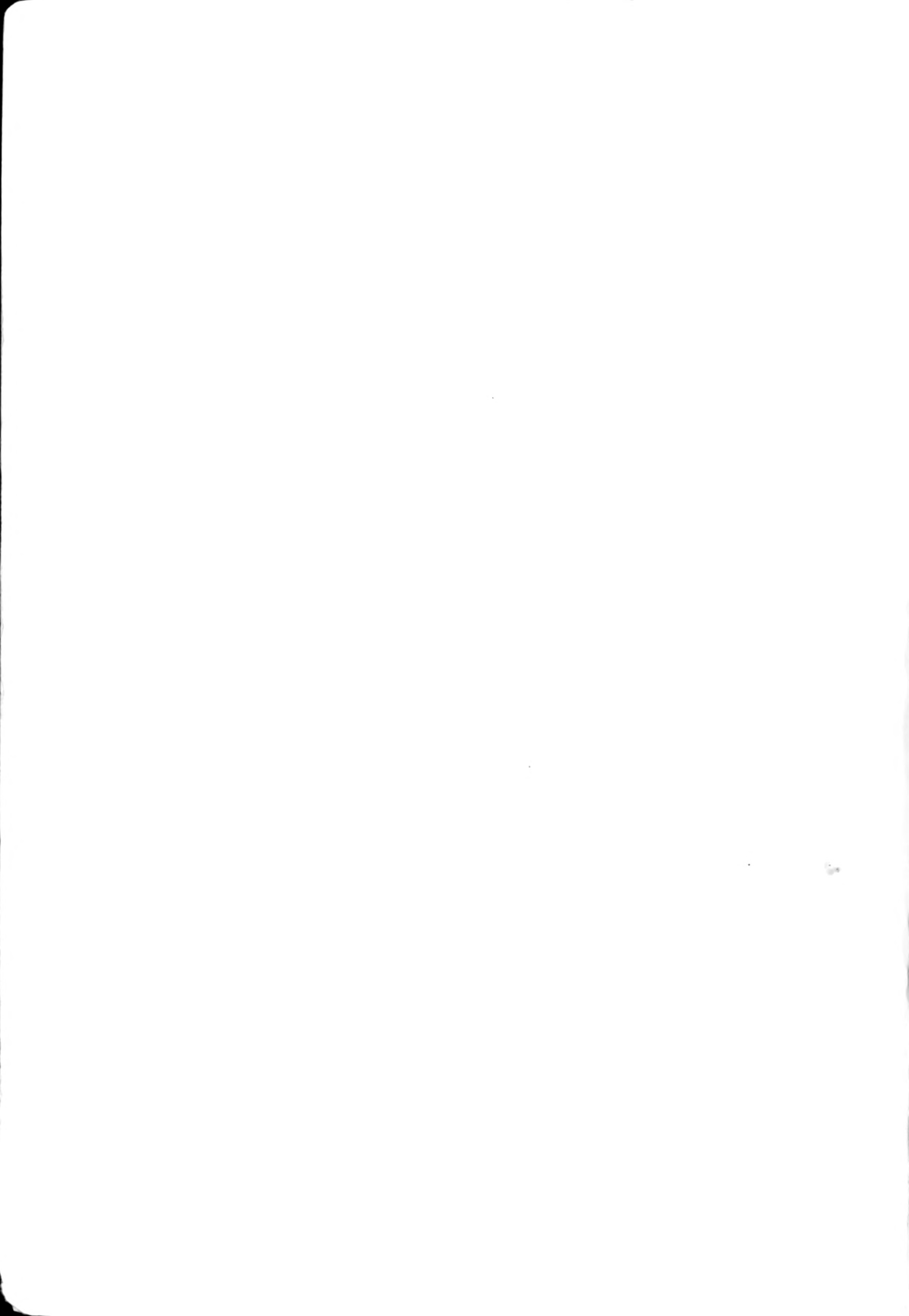
ADVANCED MOLECULAR BIOLOGY

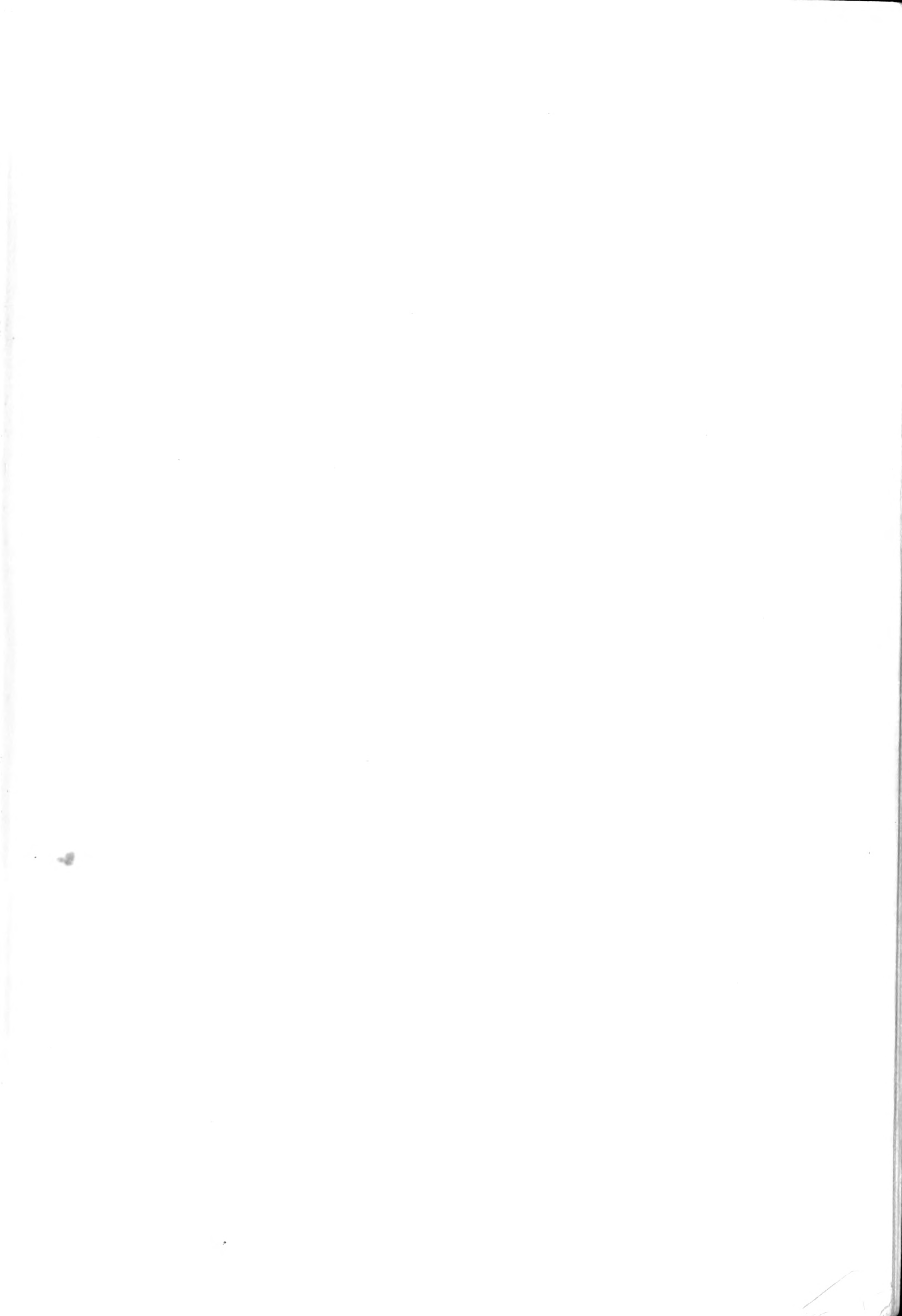
高级分子生物学要义

〔英〕R. M. 特怀曼 著
陈淳
徐沁 等译
徐晋麟 校

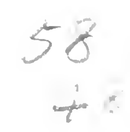


科学出版社









高级分子生物学要义

〔英〕R. M. 特怀曼 著
陈 淳 徐 沁 等 译
徐晋麟 校



科 学 出 版 社
北 京

27479

图字：01－99－0182

内 容 简 介

本书将分子生物学、遗传学的内容进行了有机的组织，分成30个主题，清晰而富有逻辑性。内容涵盖了分子生物学各重要领域，如细胞周期、细胞器和质粒、重组和复制、转录、基因表达与调控、PCR等，另外还有DNA、RNA及染色体的结构、功能、修饰、突变及修复等。本书按参考手册的风格编写，通过阐述重要理论、解释关键性术语、设置交叉参考等，向读者提供了在其他专业书中难以找到或不易把握的内容。

本书适合高年级本科生、研究生以及更高层次的人员阅读。

Richard M. Twyman
Advanced Molecular Biology: A Concise Reference
Original edition published in English under the title of *Advanced Molecular Biology*: A Concise Reference
©BIOS Scientific Publishers Limited, 1998

图书在版编目（CIP）数据

高级分子生物学要义/〔英〕R. M. 特怀曼著；陈淳，徐沁等译. -北京：科学出版社，2000. 11

书名原文：Advanced Molecular Biology: A Concise Reference

ISBN 7－03－008299－0

Ⅰ. 高… Ⅱ. ①特…②陈…③徐… Ⅲ. 分子生物－研究 Ⅳ. Q7

中国版本图书馆CIP数据核字（2000）第02615号

责任编辑：马学海 / 责任校对：陈玉凤
责任印制：刘士平 / 封面设计：王 浩

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
丽源印刷厂 印刷
科学出版社发行 各地新华书店经销

*

2000年11月第 一 版　开本：787×1092 1/16
2003年1月第三次印刷　印张：31
印数：5 501—8 500　字数：705 000

定价：55.00元

（如有印装质量问题，我社负责调换〈新欣〉）

献 给

我的父母：彼得和艾琳

我的子女：埃米莉和露西

前　言

这本书开始时是一套仓促写成的潦草讲稿，后来工整地誊写成供考试参考的笔记。由于一个朋友的过誉竟促使这本书一举出版。这位朋友借了这本笔记，读后使他一下子明白了以前某些未领会的报告，从而认为这本笔记作为一种参考资料是十分有用的。

这本书的宗旨在那时就已形成了。各章的目的提供分子生物学重要领域中的简明概论。它适合高年级本科生、研究生以及更高层次的人员。在写这本书的过程中，我遇到了很多困难并为之奋斗：当阅读一些文章、综述和其他书籍时，我和别人一样，感到要想找到要点需要翻阅很多页正文，要准确把握一个重要信息的程度也非易事。为了达到这些目的我提出了 30 个分子生物学的主题，我希望它们清晰而富有逻辑，每一主题正文限制在 10～20 页的范围内。内容又分成易操作的节段。针对每个主题在有限的篇幅中提供详细的讨论，这对那些对遗传学和分子生物学有基本了解的读者来说是一种有效的方法。这本书既不打算作为分子生物学初学者的指南，也不作为报告、综述和已出版的一些教科书的代用品，意在补充它们，并帮助读者获得关键信息。这本书从头到尾着重于定义。当关键性术语在书中首次出现时已用黑体标明，并加以详细说明。书中的图清晰是必要的，同时图的风格有意使其简明，这样不仅便于记忆，而且也易于复制。在各章节之间存在着交叉参考，是希望强调虽然书可以分成不同的主题（转录、发育、细胞周期、信号转导等）来叙述，但是所有这些过程在分子水平上都是有着重要的相互联系。在书后提供了一系列参考文献，这些文献都属于近期的一些综述或经典型文章。我希望读者感到《高级分子生物学要义》这本书是有趣的，也是有用的，但任何有关对以后编辑中的改进意见和建议都将被愉快地接受。

我想感谢很多人，没有他们的帮助和影响，这本书将不可能出版。感谢 Alison Morris，他第一个建议将仓促写成的潦草讲稿整理出版。感谢 Stuart Glover，Liz Jones，Bob Old，Steve Hunt 和 John O’Brien，他们以不同的方式在一开始就予以鼓励。非常感谢分子生物学实验室的 Steve Hunt，Mary-Anne Starkey，Nigel Unwin 和 Richard Henderson，他们一直支持我直到完稿。特别感谢那些知识比我渊博得多的 Derek Gatherer，Gavin Craig，Dylan Sweetman，Phil Gardner，James Palmer，Chris Hodgson，Sarah Lummis，Alison Morris，James Drummond，Roz Friday，他们花了很多时间阅读了每一章节，并作了评论。尤其是 Bill Wisden，他的评论和建议是极其宝贵的。最后还要感谢分子生物学 MRC 实验室的 Annette Lenton 和 BIOS 的 Rachel Offord，Lisa Mansell，Andred Bosher 和 Jonathan，没有他们的帮助这本书将不可能面世。

R. M. 特怀曼

目 录

(注：章的顺序按其英文的字母顺序排列)

缩略语表

缩略语	全称
A	腺嘌呤(碱基),腺苷酸(核酸)
AER	顶外胚层嵴
AMP,ADP,ATP	腺苷单磷酸,二磷酸,三磷酸
ANT-C	触角足复合体
AP 位点	无嘌呤或无嘧啶位点
APC	后期促进复合物
ARS	自发复制序列
ATPase	腺苷三磷酸酶
b,bp	碱基、碱基对
BAC	细菌人工染色体
BCR	B 细胞受体
BER	碱基切除修复
bHLH	碱性螺旋－环螺旋
BMP	骨形成蛋白
BX-C	双胸复合体
bZIP	碱性亮氨酸拉链
C	胞嘧啶(碱基),胞苷酸(核酸)
CAK	CDK 激活激酶
CaM	钙调蛋白
CAM	细胞黏附分子
cAMP	环 cAMP
CAP	代谢激活蛋白
CBP	CREB 因子激活蛋白
CDK	细胞周期蛋白依赖激酶
cDNA	互补 DNA
CDR	互补决定区域
cf.	比较
cGMP	环鸟苷酸单磷酸
CKI	细胞周期依赖激酶抑制物
CMV	巨细胞病毒
cpDNA	叶绿体 DNA
CREB	cAMP 反应元件结合因子
cRNA	互补 RNA
CTD	C－末端结构域
CTF	CAAT 转录因子
CTP	胞苷三磷酸
DAG	二酰甘油
Dam	DNA 腺嘌呤甲基化酶
dATP	脱氧腺苷三磷酸
Dcm	DNA 胞嘧啶甲基化酶
DEAE	二乙胺乙基
DIF	分行诱导因子
DMS	硫酸二甲酯
DNase	脱氧核酸酶
dNTP	脱氧核苷酸三磷酸
DSBR	双链断裂修复
dsDNA/RNA	双链 DNA/RNA
EGF(R)	表皮生长因子(受体)
ER	内质网
ES 细胞	胚胎干细胞
EST	表达顺序标记
FGF(R)	成纤维生长因子(受体)
FISH	荧光原位杂交
G	鸟嘌呤(碱基),鸟苷酸(核酸)
GABA	γ－氨基丁酸
GAP	GTPase 激活蛋白
GMP,GDP,GTP	鸟苷单磷酸,二磷酸,三磷酸
GNRP	鸟苷酸释放蛋白
GPCR	G 蛋白偶联受体
GTF	一般转录因子
GTPase	鸟苷酸三磷酸酶
Hfr	高频率重组
HLH	螺旋－环－螺旋
HMG	高泳动蛋白
hnRNA,hnRNP	不均一细胞核 RNA,RNP
HOM-C	同源异型复合体
HPLC	高效液相色谱
HSV	单纯疱疹病毒
HTH	螺旋－转角－螺旋

ICE	白介素 1β 转化酶
IFN	干扰素
Ig	免疫球蛋白
IL	白介素
Ins	肌醇
IRES	内部核糖体进入位点
IS	插入顺序
ITR	反转的末端重复
Jak	Janus 激酶
kb, kbp	千碱基,千碱基对
kDNA	动核 DNA
LCR	基因组控制区域
LINE	长分散排列的逆转录因子
Lod	优势对数
LTR	长末端重复
MAPK	有丝分裂激活蛋白激酶
MAR	细胞核基质相关区域
5-meC	5 – 甲基胞嘧啶
MEK	MARK/Erk 激酶
MHC	主要组织相容性复合物
MPF	有丝分裂/成熟促进因子
mRNA	信使 RNA
mtDNA	线粒体 DNA
N-CAM	神经细胞黏附分子
NAD	烟酰胺腺嘌呤二核苷酸磷酸
NCR	非编码区
NER	核酸切除修复
NMP, NDP, NTP	核苷酸单磷酸,二磷酸,三磷酸
NMR	核磁共振
NOR	细胞核组织区
OD	光密度
ORF	开放阅读框架
PAC	P1 人工染色体
PCNA	分裂细胞核抗原
PCR	聚合酶链式反应
PDE	磷酸二酯酶
PDGF(R)	血小板衍生生长因子(受体)
PEV	位置效应变异
PI(3)K	磷酸肌醇 3 激酶
PKA, PKC, PKG	蛋白激酶 A,C,G
PLA, PLB, PLC, PLD	磷酯酶 A,B,C,D
Poly(A)	多聚腺苷酸
POU	Pit – 1/Oct – 1, 2/Unc – 86 HTH 组件
PrP	朊病毒相关蛋白
PtdIns	磷脂酰肌醇
QTL	数量性状基因座
RACE	cDNA 末端快速扩增
RAPD	多态性 DNA 随机扩增
RF	复制型
RFLP	限制性片段长度多态性
RNase	核酸酶
RNP	核酸蛋白体
rRNA	核糖体 RNA
RSS	重组信号顺序
RT-PCR	反转录 PCR
RTK	酪氨酸受体激酶
SAM	S-腺苷甲硫氨酸
SAPK	压力激活蛋白激酶
SAR	核骨架附着区
SCE	姐妹染色单体交换
SCID	重度联合免疫缺陷
SDS	十二烷基硫酸钠
SH	Src 同源结构域
SINE	短分散排列的逆转录因子
snRNA	小细胞核 RNA
SR	肌浆网状组织
SRF	血清反应因子
SRP	信号识别颗粒
SSB	单链核酸结合蛋白
ssDNA/RNA	单链 DNA/RNA
SSLP	短顺序长度多态性
STAT	转录的信号传导和激活因子

STRP	短随机重复顺序多态性
STS	顺序标签位点
SV40	猴病毒 40
T	胸腺嘧啶(碱基)胸腺苷酸(核酸)
TAF	TBP 相关因子
TBP	TATA 结合蛋白
TCR	T 细胞受体
TF	转录因子
TGF	转化生长因子
Tn	细菌转座元
tRNA	转移 RNA
TSD	靶位点复制
TSE	传染性海绵状脑病
TSG	肿瘤抑制基因
U	尿嘧啶(碱基),尿苷酸(核酸)
UTP	尿苷酸三磷酸
UTR	非翻译区
V(D)J	变异,多样性,接合基因片段
VEGF(R)	血管内皮生长因子(受体)
VNTR	随机重复的可变数目
VSG	糖蛋白可变表面
VSP	极短补缀
XIC	X 失活中心
XP	着色性干皮病
YAC	酵母人工染色体
YEp	酵母基因组外质粒
ZPA	极性活性区域

如何使用本书

本书按照分子生物学的不同主题划分章节。关键性的术语用**黑体**表示,并在第一次出现时给予定义。本书提供了广泛的交叉参考,以"(参见……)"或"(参阅)"标示,读者可以从索引中进一步查找。被提示的概念术语用楷体表示。

第1章
生物的遗传和变异

基本概念和定义

- 在遗传学中,一种**性状**(character)或**特征**(characteristic)是指生物中能被描述或量度的任何生物学性质。在特定的生物群体中,性状显示两种重要的性质:**遗传**(heredity)和**变异**(variation)。这些性质可以是简单的或复杂的。大部分性状的特点是基因和环境共同作用来决定的。
- **简单的性状**表现为**不连续的变异**,即表现型能分成不同的类型,称为**性状**(trait)。这样的一些特性是符合简单性质的遗传,因此从**表现型**(phenotype)能直接推测出**基因型**(genotype),或者通过杂交分析,或者通过家谱分析(见表1.1:在传递遗传学中某些普遍使用术语的定义)。对于最简单的性状而言,表现型取决于单个基因座上的基因型。这些性状并不是仅仅只受该座位的控制,与特殊遗传背景和正常环境中的表型相对照,不同的基因型产生不同的表型。当与有性生殖的真核生物核基因组相联系时,这些性状就被描述为**孟德尔**(Mendelian)**性状**——它们符合G.孟德尔首先系统研究的不同的遗传模式。不是所有的简单性状都属于孟德尔式的。在真核生物中**非孟德尔性状**是由细胞器控制的,随之其遗传规则也不同(见细胞器基因组)。例如细菌和病毒的性状也是非孟德尔式的,由于这些生物不是二倍体,也不进行有性生殖。
- **复杂的性状**常常表现为**连续的变异**,即在两个极端表型之间变异的曲线是非常光滑的,而且表型决定于基因的数量。这种性状的遗传并不能用孟德尔法则来预见,而是用统计学方法[**生物统计学**(biometrics)]进行研究。复杂性状是由很多基因座位(**多基因学说**)来控制的。其实它们同简单性状的差别通常不是简单的相互作用的基因数目,而是环境对表型变异的影响,这种影响模糊了不同表型性状之间的差异,使得人们不能从表现型推测其基因型。

1.1 孟德尔式遗传

孟德尔式遗传原理 对于经得起遗传检验的生物(即这些生物能大量地留种和繁殖)来说,遗传的原理能经过大量的**杂交**(cross)(直接地交配)和**记录**(scoring)许多后代(的确定表型)进行研究。孟德尔将菜园中不同变异的纯种豌豆(*Pisum savitum*)进行杂交,而且将杂种植物进一步杂交,从这些结果中他发现了遗传和变异的法则。虽然他只用了一种

植物，但他的结论适用于所有有性繁殖的真核生物，包括那些不能用同样的方法来研究的动物（如人类）。对这些不能进行实验的生物，其遗传和变异是通过家谱分析来研究的（文框 1.1）。孟德尔法则可以概括如下：

(1) 性状的遗传和变异是由现在称为基因(gene)的遗传因子控制的，它成对存在。孟德尔称这些因子为形态建成因子(form-building element)。
(2) 有差异的性状是由每个基因的不同形式[不同的**等位基因**(allele)]所特有。
(3) 当两个不同的等位基因存在于同一个个体中[即在一个**异合体**(heterozygote)中]，一个性状对另一性状显示了**显性**(dominance)：其表型和一个表达的等位基因[**显性**(dominant)等位基因]相一致，而另一个等位基因作了牺牲[**隐性**(recessive)等位基因]。
(4) 基因不混合，当它们传递时，保持着独立性[**颗粒性**(particulate)]。
(5) 在减数分裂时，成对的等位基因均等分离，因此形成配子携带的每种等位基因数目相等。
(6) 每对等位基因的分离不受其他任何等位基因的影响。

表 1.1 在传递遗传学中某些普遍使用的术语定义

术语	定义
等位基因(allele)	广义，相对于一种特殊性状的一个基因的变异形式。以分子水平来说，一个基因序列的变异(参见*野生型等位基因，突变等位基因，多态性*)
性状(character)	生物能被检测或度量的生物学特征
性状类型(character mode)	性状的一般类型，如眼的颜色
性状特征(character trait)，性状(trait)，变异体(variant)	性状的特殊类型，如蓝色的眼睛
基因(gene)	广义，控制或有助于控制特殊性状的遗传因子。以分子水平而言，被表达的一个DNA(或某些病毒中的DNA)片段，即在细胞中用于合成一个或多个具有特殊功能产物的DNA(或RNA)片段(参见*基因、顺反子、基因表达*)
(基因)座(locus)	一个基因(或其他标记或界标)在一条染色体上或在物理图或在遗传图上的位置。这是一个有用的术语，因为它使得人们讨论基因时可不考虑基因型或接合性
遗传的(genetic)	关于基因的，由基因的核苷酸序列产生的性状、遗传和变异(对照*外遗传的、环境的*)
基因型(genotype)	一个个体的遗传性质，常涉及特定座位上的等位基因的特殊组合
半合子的(hemizygous)	在二倍体细胞中含有一个等位基因，常涉及*性连锁基因*(参阅)
遗传性的(hereditary)	从亲代传递到后代。其范围比“遗传的”这个术语更广，包括遗传的继承(核苷酸序列的遗传)，也包括外遗传的继承(DNA结构中的信息的遗传)以及细胞质遗传或细胞分裂时的膜成分
杂合的(heterozygous)	在一个特定座位上含有不同的等位基因
纯合的(homozygous)	在一个特定座位上含有相同的等位基因
表现型(phenotype)	一个个体的外在性质，常涉及特殊性状的性质
多效性(pleiotropic)	同时影响到一个以上的性状
变异(variation)	在特定群体中，一个特异性状的变化。变异可以是连续的或不连续的
接合性(zygosity)	在一个基因座位上等位基因的性质——纯合、杂合或半合

为了使基因和等位基因结构和功能的定义更准确，见“基因”和“突变与选择”章节。

1.2 一个基因座上的分离

在一个基因座上的杂交 孟德尔原理中的5项可能是从**单点杂交**(单一因子杂交)中推论而得出的,为了便于研究将单个基因座位分出,在不同纯种品系之间进行杂交产生杂种后代并建立了**显性原理**(图1.1)。一个纯系品种当自交或杂交时对其特殊性状是稳定的,由此可见纯系只含有一种类型的等位基因,即所有个体在特定基因座位上是纯合的。两个不同纯系之间的杂交产生的杂种后代,它们所有个体都是杂合的,携带着来自各纯系的一个等位基因。这是**第一代子代**(F_1代)。孟德尔在其杂交实验中出现了F_1代杂种的表现型和一个亲本相同,即一个性状对另一个性状是显性的。

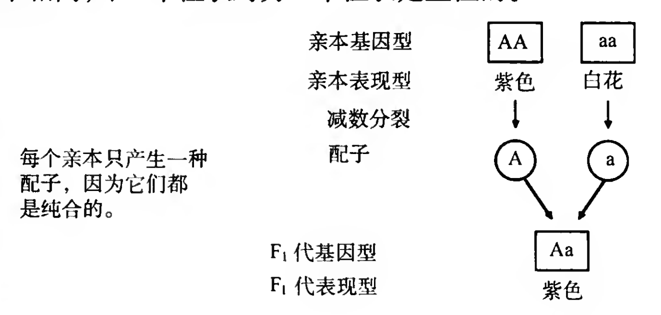


图1.1 纯系之间的杂交。产生了杂种F_1代并建立了显性原理。其中A等位基因在纯合中形成了特异的紫色花,它对于a等位基因是显性的,a等位基因在纯合时形成特异的白色花。花色的基因座位被发现在豌豆这种植物的1号染色体上,人们认为它编码了一种和色素产生有关的酶;a等位基因被认为是无效的。

回交(backcross,涉及到子代和一个亲本的杂交)能确定F_1代是杂合的。若将F_1代与带有隐性等位基因的纯合亲本进行杂交,回交后代中表型比例为1:1的话,那么可确定F_1的基因型(图1.2)。这种类型的分析表明和**测验种群**(test stock,所研究的各个基因座都带有隐性等位基因)有关的遗传杂交能决定未知的基因型,而相同的原理可被用于遗传作图(参阅)。在F_2代中隐性表型(即白花)的存在可以确定在基因传递中,成对的等位基因仍保持着颗粒状,在杂种产生的表型中它们既未被取代,也没有混合。

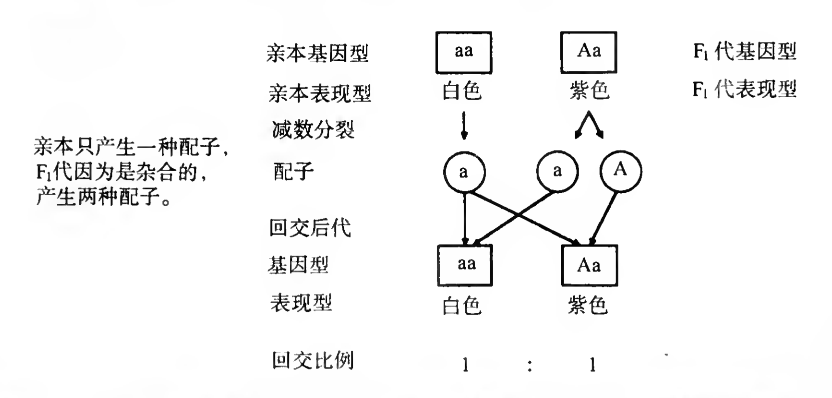


图1.2 F_1杂种与其隐性亲本之间的回交。由于隐性表型出现在后代中,这种杂交表明了F_1代是杂合的,即隐性的等位基因仍作为一个独立的单位存在。

F_1 **自交**(self-cross,自体受精)或者(有的生物不能自交)F_1个体间的**互交**(intercross)

可以称为**单基因杂种杂交**(monohybrid cross),这是由于参与者在一个特定座位上是杂合的。这种杂交论证了均等分离的原理,此已被叫做**孟德尔第一定律**(Mendel's First Law)。在随后的**子二代**(F_2 代)中表型的比例是 3:1(图 1.3)。这就称为**单基因杂种比例**(monohybrid ratio),并预期在减数分裂中若等位基因均等分离的话才会发生(产生相等数目的带有二个可能等位基因中任一个配子)。

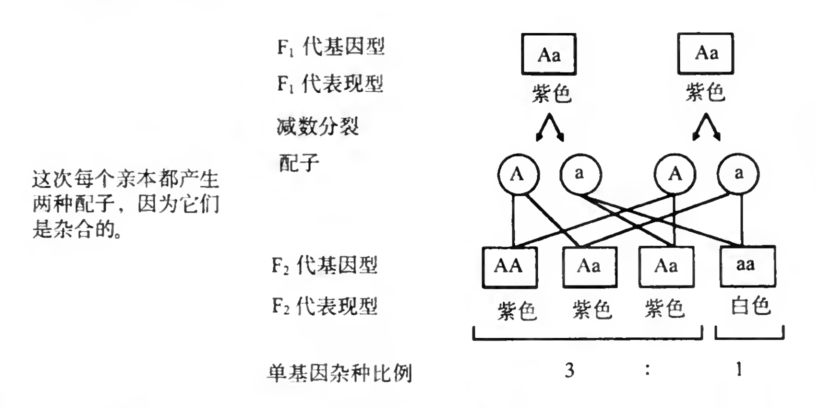


图 1.3 单基因杂种杂交。3:1 的单基因杂种比表明在减数分裂时等位基因均等分离。图中所用交配图谱显示在受精时所有可能的配子组合。

单基因杂种比例的偏差 孟德尔遗传模式广泛地适用很多性状,但从单基因杂种分离比的讨论中出现了特异的偏差。这些性状的分析论证了孟德尔的 6 项原理中有一项可能被打破,事实表明孟德尔的运气很好,在他所选的 7 项性状中都未遇到任何下面将讨论的一些麻烦。从未被打破过的法则就是颗粒遗传的原理——基因并不混合,当它们被传递时是保持作为独立的单位。

亲代的不相等性 从孟德尔实验中得出的主要结论之一是对于各个基因座位来说,一个等位基因是从单个亲本中获得的,即双亲对合子的贡献是相等的。**互交**(基因型为 A 的雄性和基因型为 B 的雌性杂交及相反的情况)的结果是相等的,此是谱系中**常染色体遗传**的基础(文框 1.1)。亲本的相等性反映了这样的事实,二倍体真核生物常有两套染色体,从每个亲本中获得一套。在大部分情况下,两个亲本贡献相同数量的染色体,而且活性也是相同的。有两个重要的例外:性连锁遗传(由于结构的半合子性)和单等位基因表达(由于功能的半合子性)。

性连锁遗传 性染色体(参阅)控制了动物的性别决定(参阅)且在两性间的分布是非对称的(参见常染色体)。有一种**同配性别**(homogametic sex),它们具有一对相同的性染色体,这样就只产生一种类型的配子,而**异配性别**(heterogametic sex)具有一对不同的性染色体,或是只有单条不成对的性染色体,这样可产生两种类型的配子。

在哺乳动物中雌性是同配的:她们带有两个拷贝的 X-染色体,而雄性常有一条 X-染色体和一条 Y-染色体,因而是异配的。在 X-和 Y-染色体之间有两个短的同源区,**主要和次要拟常染色体区**(major and minor pseudoautosomal region),在减数分裂中它们促进配对。主要的拟常染色体区是专性交换位点,而位于此处的基因以正常的**常染色体式样**

遗传(拟常染色体遗传,pseudoautosomal inheritance)。另一些基因被描述为性连锁的,因为它们的表达取决于性染色体的分布如何。

在杂交中,**X-连锁基因**(X-linked genes)能被鉴别出来,这是由于和单基因杂种互交的结果不同(图 1.4)。若显性的等位基因是由雌性携带,则可获得正常的孟德尔比,但若显性的等位基因由雄性携带,则在 F_1 和 F_2 代二者中会产生特异偏差,此是因为雄性是半合子之故。在另一种情况下,X-连锁基因显示出表型的性别特异性,反之常染色体传递的性状其分离比是性别非依赖性的。此外,由于雄性遗传的 X-染色体来自他们的母亲,而 X-染色体又只传递给他们的女儿,这样性别表型的关系在每代中交替变化,这种现象称为**绞花式遗传**(criss-cross inheritance)。此是用于鉴别人类谱系(文框 1.1)中 X-连锁遗传的主要特征。

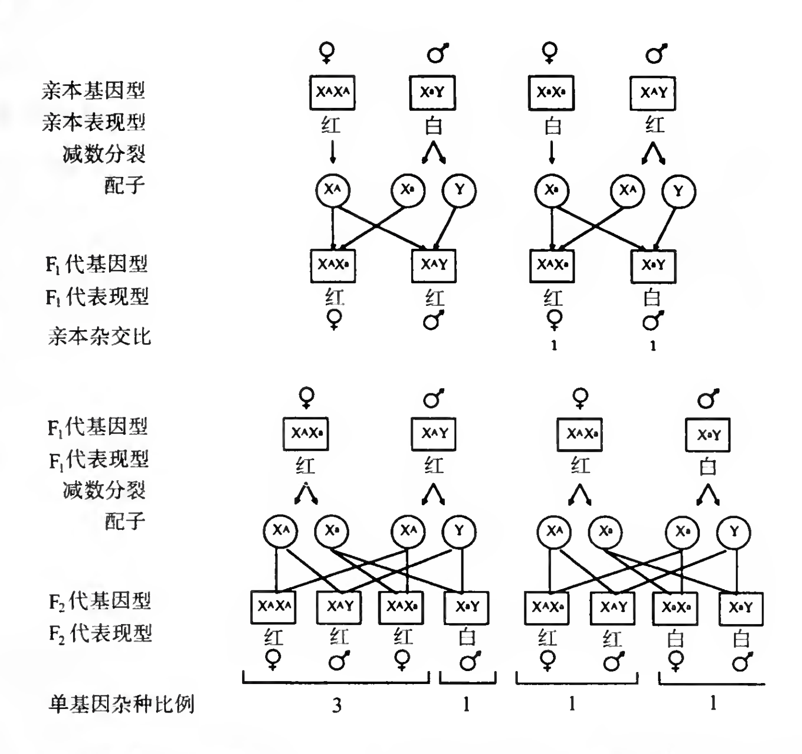


图 1.4 X-连锁遗传。由于雄性是半合子,互交的结果是不同的。分离比连锁于性别比例,在性别特异表型中结果雄性总是将其 X-连锁的等位基因传给他的女儿。

在杂交和谱系中,**Y-连锁基因**可以被鉴别,因为 Y 染色体控制的性状仅在雄性中表达,且只通过雄性进行传递(**限雄遗传**,holandric)。很少有 Y-连锁性状在人类中被鉴别出。

单等位基因表达 某些来自双亲的常染色体基因只有一个等位基因是有活性的。此就称为**单等位基因表达**(monoallelic expression),而此座位是功能性的。而不是结构性的半合子。有二种类型单等位基因表达:**亲代印迹**(parental imprinting),由一个亲本遗传的基因特异性被阻遏,而**随机失活**(random inactivation)是亲代等位基因被随机选择性的抑制。在哺乳动物中有两种随机失活的方式——**X-染色体的失活**和免疫球蛋白基因表达的**等位排斥**(al-

lelic exclusion)。单等位基因表达在这一章中将不再进一步讨论——在进一步讨论亲本印迹和X-染色体失活中见DNA甲基化和外遗传调节以及讨论等位排斥中的重组。

母体效应和母体遗传 在各种其他情况下正反交的结果是不同的。一个例子是母体效应,在**母体效应**(maternal effect)中,个体的表型完全取决于母体的基因型,而和父本的基因型无关。母体效应是在发育早期发挥功能的那些基因中观察到的,它反映了这样的事实,母体的一些基因产物进入了卵中,这些产物是用母体的基因组在母体细胞中合成的。显示母体效应的基因实际上是以正常的孟德尔方式遗传的,但相应的表型在子代中观察不到(见图6.1),这取决于胚胎母体祖父母的(相等)贡献。在祖父母这一代中进行正反交结果胚胎的表型应是相同的。在这种方法中,母体效应不同于*母体遗传*(参阅),一些非孟德尔式遗传的基因只能通过母系来传递,因为这些基因位于胞质中的细胞器基因组内,而不在核基因组中。母体遗传的基因并不特异地和发育相关(如它们的表达贯穿了个体生命的始终),且在任何后代中都没有雄性的贡献。这样互交在所有后代中结果都是不同的。关于母体遗传和其他形式的非孟德尔遗传我们将进一步讨论。见细胞器基因组。

等位基因的变异和相互作用 孟德尔描述的性状有两种类型,即它们是**二对等位基因的**(diallelic)。但有很多性状存在大量的**等位变异**(allelic variation)。例如人类的ABO血型座位有3个生理学特性的等位基因,在丁香和烟草的自交不育的极端例子中,在特异的群体中可以检测到超过200个不同的等位基因。等位基因变异的观察依赖于表现型被确定的水平。在分子水平的差异常比形态学水平的差异更大,这是由于很多等位基因被鉴别出序列的变异或蛋白质的多态性(见突变和选择)。它们对于形态学的表型的影响是中性的;这些就叫做**同等位基因**(isoallele)。无论怎样在一个群体中对于特殊位点而言很多等位基因可能是有差异的,在任何时候,在同一个二倍体个体中只存在两个等位基因。形态学特异的等位基因可能常列为显性的,因而称为**等位系列**(allelic series)。

在每一组孟德尔杂交中,一个与性状相关的等位基因对于另一个是完全显性,因此杂合体的表型和显性纯合体是相同的。从生化水平来说**完全显性**(complete dominance)常反映了存在一个(隐性)*无效的等位基因*(null allele)(参阅),总体上它由一个(显性)正常功能的等位基因来补偿;此常发生在编码一种酶的座位上,由于大部分酶在低浓度时是有活性的——豌豆中紫色花瓣色素的酶就是一个例子,但在其他植物中不一定存在这种情况,有时会导致不完全显性。存在一些**显性关系**(dominance relationship)和另一些**等位相互作用**(allelic interaction),但每种都有特异的生化基础,这些在表1.2中阐明。

显性的概念常常适用于等位基因,但显性是性状本身的一种性质,而不是等位基因控制它们[仅在*副突变*的情况下(参阅)等位基因本身发生了可遗传的改变]。显性也取决于表型的观察:镰刀细胞特征是部分显性的疾病,因为就正常细胞和镰形细胞中变异β-珠蛋白产物而言,其等位基因在杂合子中的作用是明显的,(虽然是在极端的情况下),但从蛋白质的水平上观察电泳胶上移动的带,变异的β-珠蛋白与正常蛋白是共显性(即两种性状同时都能观察到)。

分离比偏差 均等分离的原理是孟德尔法则中较为重要的准则之一,且是从观察单基因杂交后代中亲本所贡献的等位基因均等分离的现象中获得的。但存在抑制均等分离的各种途径,结果使孟德尔式比例发生变形——如在后代中倾向于特殊等位基因的恢复。

这种机制分为两种主要类型:受精前作用和受精后作用。

表 1.2 显性关系和另一些等位相互作用(在单个座位上的相互作用)及其生化基础和例子

等位相互作用	叙　述
完全显性	显性等位基因完全屏蔽了隐性等位基因的作用。杂合子的表型和显性等位基因的纯合子相同。单基因杂种的比为 3:1。这是由孟德尔描述的典型的显性作用,常发生的隐性等位基因是无效基因。例子包括豌豆中的紫色色素和哺乳动物中的囊性纤维化(cystic fibrosis)。失去一个编码色素酶的等位基因或跨膜受体将会被第二个野生型等位基因所补偿。另一种可能是显性等位基因是无效的(参见*显性失活*,dominant negative),例如在先天性丘结肠症(Hirschsprung's disease)中通过显性失活失去了 c-RET Tyr 激酶活性所致——这一突变使野生型酶的顺序变成为失活的异二聚体。
无显性和部分显性	两个等位基因中的任一个对另一个基因都不是完全显性。杂合的表型居于两个纯合体之间,单基因杂种比为 1:2:1 如果杂合体表型准确地表现为两个纯合体之间的中间状态,那么就是无显性。如果杂合体表现偏向一个纯合体,那么就是部分显性。在这种显性关系发生处具有两个等位基因产物之间的竞争。[例如在镰形细胞特征(sickle-cell trait)中,不同形态的 β-血红蛋白对低氧张力的反应不同]或者是有一个基因座位是*单倍不足*(haploinsufficient)(例如在 I 型瓦尔登伯格综合征(Waardenburg syndrome)是由于 PAX3 蛋白的合成中产物减少了 50%)。
超显性和显性不全	杂合体的表现位于显性纯合体所描述的范围以外。若杂合体的表现比二种纯合体中任一种都要超出一些的话,这个座位就显示了超显性(overdominance);若少一些,是显性不全(underdominance)。单基因杂种比为 1:2:1。这种关系发生在特殊等位基因产物之间的协同作用和反作用。当考虑到导致**杂种优势**[hybrid vigor (heterosis)]的多座位复合效应时常观察到超显性,这是由于在很多座位的杂合增加了适应性;或者**近交衰退**(inbreeding depression),这是由于很多有害的等位基因被纯合而减少适应性。
共显性	一对等位基因的相关表型都可以独立表达。当在等位基因之间不存在竞争时发生共显性(codominance),例如在 ABO 血型系统中等位基因 A 和 B 控制合成红细胞膜表面的不同糖蛋白,A 和 B 二者对于无效等位基因 O(即蛋白保持非糖基化)来说都是显性的。然而若一个个体带有 A 和 B 两个等位基因。则两种糖蛋白的分子都存在,且产生了 AB 血型。其单基因杂种比为 1:2:1。
拟显性(pseudodominance)	一个等位基因呈现显性是由于该座位是半合子的。这适用于异配性别中的性连锁座位,如在雄性哺乳动物(参见*性连锁*)和带有染色体缺失或染色体丢失的个体(染色体突变)。
副突变(paramutation)	在杂合状态中发生的等位相互作用(allelic interaction)时,有一个等位基因引起了短暂的可遗传性,而另一个基因产生外遗传改变,这个过程常常涉及重复 DNA 的甲基化。这仅是 DNA 本身作为靶的一种等位相互作用的例子。进一步讨论见 DNA 甲基化。
等位互补	两个没有功能的表型相对于野生型等位基因来说都是隐性的,经重组它们能产生有功能的基因产物,这是由于它们互相补充了彼此的缺陷。等位基因互补(allelic complementation)的主要例子是大肠杆菌 β-半乳糖苷酶表达中的 α 互补(参见*重组体选择*)。
反式感应(trans-sensing)	在生物体的同源染色体上的联会-依赖性等位基因之间的相互作用,仅发生在同源染色体联合时,甚至是有丝分裂细胞(如果蝇)或在其中偶尔发生的结合。例子包括联合调控(transvection),进一步讨论参见基因表达和调节。

发生在受精前**分离偏差**(segregation distortion)(如带有等位基因不同成员的配子存在着不均衡性)也被称为**减数分裂驱动**(meiotic drive)。有两种不同类型的驱动机制优先发生在不同的性别中。**基因驱动**(genic drive)通常发生在雄性中并和特殊单倍体精子选择性失活有关。两个座位涉及到这一系统的类型,反式作用**驱动**或**偏差**以及顺式作用**靶**。在果蝇的 *SD*(分离失调基因,segregation distorter)系统中靶等位基因是重复 DNA 元件,其拷贝数是与偏差的比例相对应的。此驱动座位编码的产物被认为作用于靶等位基因上,从而打乱了染色质的结构,导致配子的机能障碍,因此 DNA 的结构调节可能广泛地用于配子失活机制。基因驱动在雌性中并不是普遍的,这是因为产生的雌性配子远比雄性配子少得多,它们可能通过大规模的配子失活将处于选择不利的境地。在雌性中的驱动常比雄性中发生得早,这是通过称为**染色体驱动**(chromosomal drive)的过程,在此过程中减数分裂产生的二价体的特点致使它采用特殊的纺锤体中的方向,这样偏向分离成卵细胞或是极体(后者被摒除)。在雄性中将不进行染色体驱动,因减数分裂的产物是均等的。

受精后产生偏差反映了带有不同基因型的合子具有不同的生命力。最为极端的例子是偏差导致在总的基因型中缺乏特殊的基因型,表明存在**致死等位基因**(lethal allele)(当它们表达时使个体死亡),它们的作用在发育早期最明显。显性致死基因的存在使得只有一种基因型即纯合隐性得以幸存。存在隐性致死基因的后代其显性纯合和杂合的分离比为 2:1,因纯合隐性的类型是不能存活。致死等位基因通常表现为主要基因产物功能丧失(参阅);这样渗漏致死等位基因可能是**亚致死的**(sublethal)(参见外显率、表现度、渗漏突变)。

外显率和表现度　外现率和表现度是常用于描述简单性状的遗传背景、遗传环境和遗传噪音的非特异效应(文框 1.2),**外显率**(penetrance)是描述一种特殊基因型显示出相应表型的个体比率。当基因型和表型 100% 符合时发生完全外显。**表现度**(expressivity)反映了特殊基因型表达的程度,即表型可以用一种严格的术语来权衡,最强的作用具有最大的表现度,不完全外显和可变的表现度常使人类谱系难以解释,这和少数个体有关。不完全外显和可变的表现度其作用特点是使谱系不再能可靠地从表型来确定其基因型,此特点被描述为复杂性(见下文)。

1.3 在两个座位上的分离

在两个座位上的杂交　孟德尔最终的假定表达为**自由组合原理**(principle of independent assortment),此可以从**双因子杂交**(two-factor cross)(通过一次杂交同时研究两个座位)中推导出。具有二个不同性状的品系经杂交产生的 F_1 代都为双因子杂种(在二个座位上杂合)。若是自交或互交(双因子杂种杂交)产生的 F_2 代可能有 4 种表型,分离比为 9:3:3:1。这被称为**双因子杂种比例**(dihybrid ratio),并在图 1.5 中显示其来源。若一对等位基因的分离不影响另一对的话,即两对等位基因在减数分裂中是自由组合的话,只产生双因子杂种比例。自由组合原理被称为孟德尔第二定律(Mendel's second law)。

在亲本杂交中，每个亲本只产生一种配子，这是因为它们在两个座位上都是纯合的。

亲本基因型 RRII rrii
亲本表现型 黄色 光滑 绿色 皱缩
减数分裂
配子 RI ri
F_1 代基因型 RrIi
F_1 代表现型 黄色/光滑

…但在双因子杂交中，F_1代可以产生四种配子，因为它在两个座位上都是杂合的。

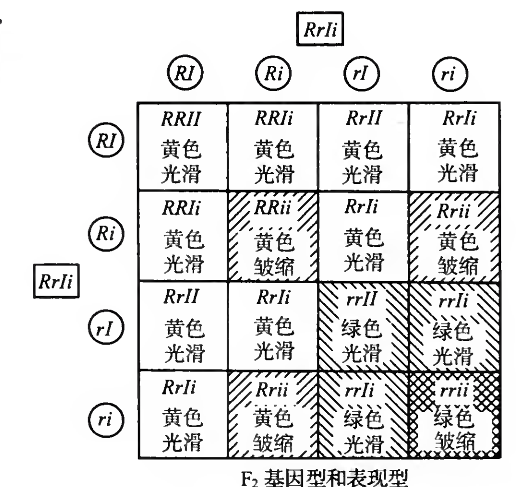


图 1.5 自由组合原理。两个亲代品系选择含两个不同性状的纯系(黄色对绿色，光滑对皱缩，这些性状由 1 号染色体的 I 座位和 7 号染色体的 R 座位编码)亲代杂交产生相同的 F_1 代中双因子杂种在每座位上都呈现了显性性状(光滑黄色的种子)。每一个 F_1 个体都产生四种配子，在后代中它们可以组合成 9 种不同的基因型和 4 种表型。若每一对等位基因彼此独立分离那么应得到 9:3:3:1 的分离比。格子表示双因子杂种中的交配信息称为**庞尼特方格**(Punnet square)，它比同时分析两座位的杂交图更为有用。

然而，当两个基因座位在染色体上彼此靠得很近的话，偶联的等位基因倾向于一道分离。这 9:3:3:1 的双因子杂种比被另一种比例所取代，在这种比例中两种表型多于理论比，另两种类型少于理论比。比理论数多的类型表现为等位基因的亲本组合，即在同一条染色体上基因偶联。同时比理论数少的类型是在减数分裂时一对染色体之间交换而产生的等位基因重组合。这种现象叫**连锁**(linkage)，并可用于真核生物基因组的作用：见重组和基因组与作图。

非等位基因相互作用 连锁打破了这种杂种比例，这是由于抑制了独立分配。若两个座位是功能性相互依赖，即两个座位控制同一性状，在不连锁的情况下双因子杂种比例也可被修饰。各种类型的**非等位基因相互作用**(nonallelic interaction)的发生产生与正常 9:3:3:1双因子杂种比例的特殊偏差(表 1.3，图 1.6)。

表 1.3 某些不同类型的非等位基因的相互作用，产生被修饰的杂种比例，这种相互作用是遗传背景的基础

相互作用	描　述
无	两个分离的座位不连锁且控制不同的性状。双因子杂种比例为 9:3:3:1。
加性效应（additive effect）	两个分离的座位以累加的形式对同一性状起作用，等位基因一般作为正的或中性的作用，最终表型决定于所有加性座位总的贡献。如两个座位是独立的(参见*数量遗传*)，那么双因子杂种比应为 9:3:3:1。
基因的互补作用	有种情况是两个基因座位上的特殊等位基因负责一个特殊的表型。若二个座位中任何一个座位缺乏有功能的等位基因，则这种表型将不会产生。若二个座位需显性等位基因的话，双因子杂种比将是 9:7；若两个座位需隐性等位基因的话，比例为 1:15；若一个座位需显性，另一个座位需隐性的话，比例为 3:13。
上位（epistasis）	一个座位上的等位基因抑制或掩盖第二个座位上基因表达的现象，这第二个座位称为**下位**(hypostatic)。上位可以是显性的或者隐性的，即在上位座位需一个或二个等位基因才能发挥作用。双因子杂种比例分别为 12:3:1 和 9:3:4。上位常在遗传途径或(染色体)级系中可以观察到。在这些途径或级系中，基因的作用有一定的顺序，早期作用的基因对于晚期作用的基因是起上位作用的。**表型修饰**(phenotypic modification)除**修饰等位基因**而不是掩盖下游基因的表型之外和上位是相似的。
重复/剂量	由于重复而产生的相同等位基因可能以表面上像是等位基因的方式进行相互作用。但它们可以通过重组而分离，表明它们发生在不同的座位上。这种相互作用称为**拟等位基因**(pseudoallelic)，其双因子杂种比例应为 15:1。多次重复常可在转基因生物中见到，在那儿很多转基因的拷贝汇集。转基因动物可能显示出**基因表达依赖拷贝数**(copy-number-dependent gene expression)(即基因产物的水平和拷贝数一致)或者重复拷贝之间相互作用导致*同源性依赖的基因的沉默*(参阅)(homology dependent gene silencing)。
抑制	在一个座位上的等位基因[**抑制等位基因**(suppressor allelt)]以某种方式通过补偿其效应来阻止第二个座位上的特异(通常是有害的)等位基因的表达。这个术语常用于描述一种相互作用，在此相互作用中，一个突变补偿第二个不同座位上的突变使野生型表型得以恢复(参见*第二位点突变*)。抑制可能是显性的或隐性的，这取决于这个抑制座位上是需要一个还是二个等位基因。其双因子杂种比应分别为 12:3:1 和 9:3:4。上位不同于抑制在于前者是基因特异的，而不是等位特异的(即上位阻止了所有等位基因在特殊座位上的表达)，后者是基因特异和等位基因特异的(即抑制是在第二个座位上补偿了特殊等位基因的作用)。
协同作用、增强作用	除了**增强子等位基因**(enhancer allele)特异增强第二个座位的突变效应以外，和抑制的机制相同。它可以取代抑制。和抑制等位基因一样，增强子等位基因可以是显性和隐性的。协同增强作用不同于相加效应，因增强子座位是不能对单独被增强的靶座位的相关表型作出贡献的，即若靶等位基因并不存在，那么增强子座位对表型无效应。增强子等位基因不应和*增强子*(enhancer)调节元件相混淆(参阅)。

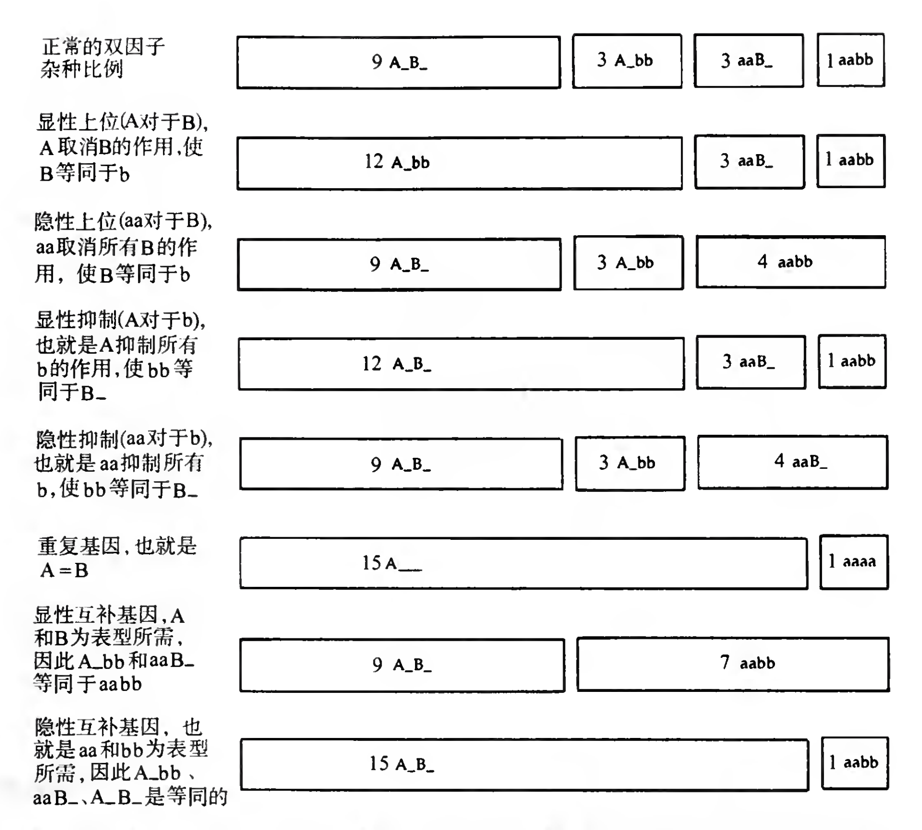


图 1.6 非等位基因的相互作用对孟德尔双因子杂交比的影响。两个假设座位 A 和 B,每个座位由一对等位基因构成,其中一个用大写字母,表示对另一个是完全显性的。对于正常的自由组合来说,4 种表型相应的一般基因型为 A_B_,A_bb,aaB_和 aabb,它们是按一定比例(9:3:3:1)。不同类型非等位基因的相互作用是通过改变与特殊等位基因相一致的表型来调整比例得以显示的。

1.4 数量遗传

复杂性状的类型 很多性状表现出连续的变异,即表型是以数量的术语来表示且不能列为不连续的性状。表型常是围绕中值的正态分布。这种**数量性状**(quantitative character)是以一种复杂的方式遗传的:不能从表型来推测其基因型,也不能用简单的遗传法则来预期杂交的结果。这种性状的遗传是用统计学方法(生物统计,biometrics)来研究的。然而有些表面上表现为孟德尔式的性状也是属于**数量遗传**(quantitative inheritance)的。数量遗传的规律性包括三类性状(图 1.7)。**连续性状**(continuous character)表示真正的连续的表型变异(即在两个不同的表型之间没有明显的界线)。**数量性状**(meristic character)与连续性状十分相似,但这种性状本身内在的性质要求表型列为不连续的种类,通常这是由于这些表型的数值要通过计数来决定(今后这种性状也可称为连续性状)。最后是**临界(二叉)性状**(threshold/dichotomous characters)有两种表型——在一定的域值内可能存在或是缺乏某种性状。这种性状非常象孟德尔的二对等位基因的性状,但在这种情况中一个不明显的数量机制控制着显示表型的能力,一旦超出了某一触发水平表型就会显

示出来。

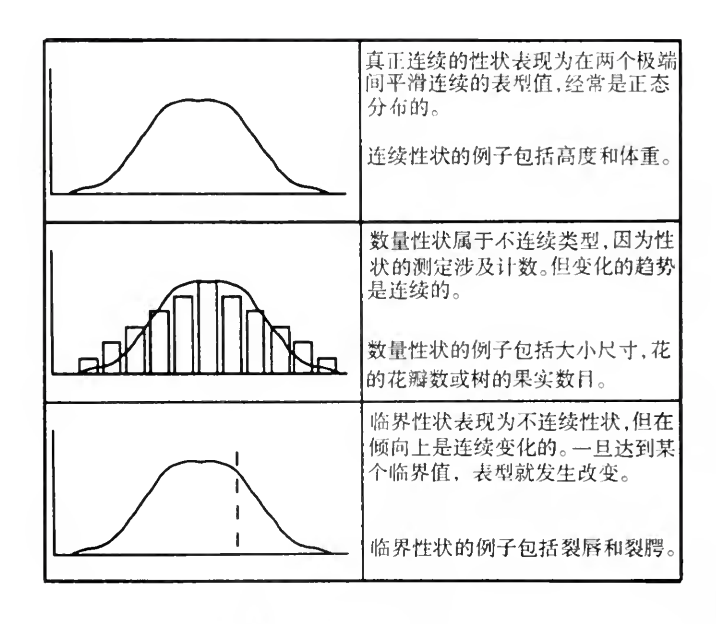


图 1.7　数量性状的类型。

数量遗传的多基因学说　　由于数量性状不适合于孟德尔分析，原来认为控制这些性状的因子基本上不同于控制孟德尔性状的基因。

多基因学说(polygenic theory)首先建立在孟德尔性状和数量性状传递的共同基础上。此学术提出数量性状可能由大量的分离的基因座位[**多基因**(polygenes)]控制的，每个座位是以一种小的但累加的方式对性状作出贡献。以这个模型的最简单的形式，每个贡献座位是双等位基因的，其中一个等位基因对表型作出贡献，与此同时另一个不起作用。每个等位基因的结果是相同的，这样表型就取决于作出贡献的等位基因的数目，即表型是由多个等位基因总计的结果来决定。当分离的基因座位数增加时表型的数值也增加。其分布为正态曲线(图 1.8)，这是由于在表型数值的中间区域有很多不同的基因型将产生相同的表型[**座位异质性**(locus heterogeneity)]。

对于数量遗传同时提出了十分相似的模型，真实的群体预期可能会遇到前一节已讨论的某些难题。这样观察一些遗传的模式由于以下的原因可能更为复杂。

(1) 每个基因对表型的相对贡献可能不同——某些座位可能具有强的作用，而另一些仅有弱的作用(这些被分别称为**主要基因**(major gene)和**次要基因**(minor gene)；
(2) 由于连锁分离的特点可能变得复杂化，在同一条染色上的等位基因其重组频率将会增加；
(3) 在某些对表型有作用的座位上可能具有两个以上的等位基因，它们可能具有不同的能力；
(4) 在等位基因之间可能有显性关系；
(5) 在有作用的座位之间除了相加效应以外还可能存在非等位基因的相互作用；
(6) 在自然群体中，不同的等位基因将以不同的频率存在，因此某些基因型比另一些稀少

的基因型相对要常见。

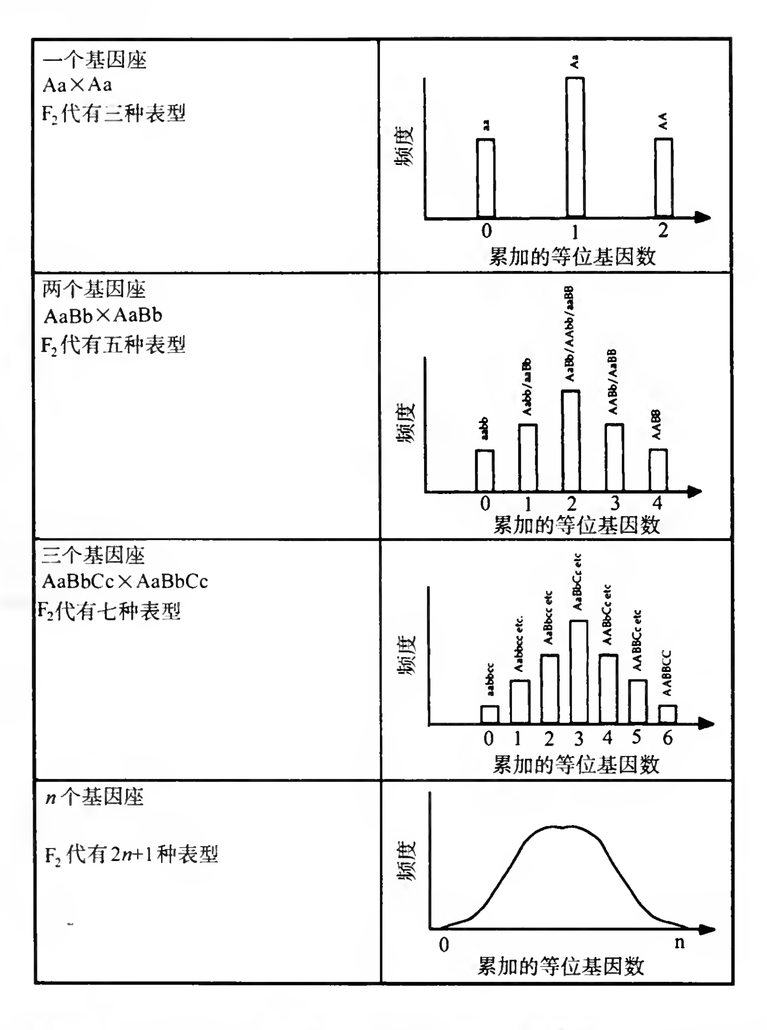


图 1.8 数量遗传的多基因模型。当影响一特定性状的加性座位数量增加时，F_2代中表型的数也增加。每个座位上两个相加的等位基因能对表型贡献有一个极大值，因此 n 个座位就有 $2n+1$ 表型。在中间区域，有很多不同的基因型产生相同表型。此分布呈现正态曲线，且表型变异是连续的。

环境影响和反应规范　尽管多基因模式简单有趣而应用广泛，但缺乏最终的证据证明数量性状是由于多基因控制的。人们系统探索那些控制数量性状的基因，少数主要基因和数目不定的次要基因常被鉴别出来，表明有少数座位就足够了。对于人类中以一种复杂方式遗传的疾病，其主要基因称为**主要的易感座位**(major susceptibility locus)(参见数量性状座位，QTL 作图)。

事实上，连续的变异并不需要大量的基因座位，因为所有生物性状象受到基因型影响一样，受到一定程度的**环境**影响。生物所处的环境与基因型相互作用而产生表型。这样若单一的基因型与环境区域相接触，则被观察到的表型范围描述为**反应范围**(norm of reaction)。此可解释在等基因群体中(isogenic populations)(群体中各个体具有相同的基因

型)当所有个体不处在相同环境时,则有很多的表型变异。[另外,等基因的个体非环境的方式也可能不同,包括存在着体细胞突变(在脊椎动物中以体细胞重组的方式重排胚系的免疫球蛋白和T细胞受体基因),和在雌性哺乳动物中具有活性和失活的X-染色体的分布。]可能由环境形成的表型的程度被描述为**表型可塑性**(phenotypic plasticity)。

简单表型的存在是由于每种基因型平均表型之间的差异要大于各种基因型的反应规范(换而言之,即基因型之间的变异要比基因型内的变异大得多)。对于连续性状反而言之也是正确的:每种基因型平均表型之间的差异比各基因型的反应规范更为相似,因此后者是重叠的。此重叠的意思是指从表型不能预见其基因型,而孟德尔分析对数量性状也无能为力——若性状是由单个分散的座位控制的话,则基因型可以预见,也可用孟德尔分析。但对于多基因性状而言,当座位的数目增加时表型的数目变得很大,而它们之间的差别变得更小。这样较少的环境影响是彼此界线完全变得模糊所需的(图1.9)。通过控制环境,使得反应规范小,从而受少数几个基因座控制的连续性状被分解成不连续的性状,且它们的传递可以用孟德尔遗传来分析。

相对地说少数情况是属于纯粹的孟德尔遗传、纯粹的多基因遗传或完全由环境决定的。大部分性状是属于这三者之间。孟德尔性况可以看作三角形的顶端,既不受遗传背景的影响,也不受环境的影响。当其他基因和环境的供量(contribution)增加时,性状将开始显示外显率不完全,表现度不稳定并将最终变成为数量遗传(图1.9)。

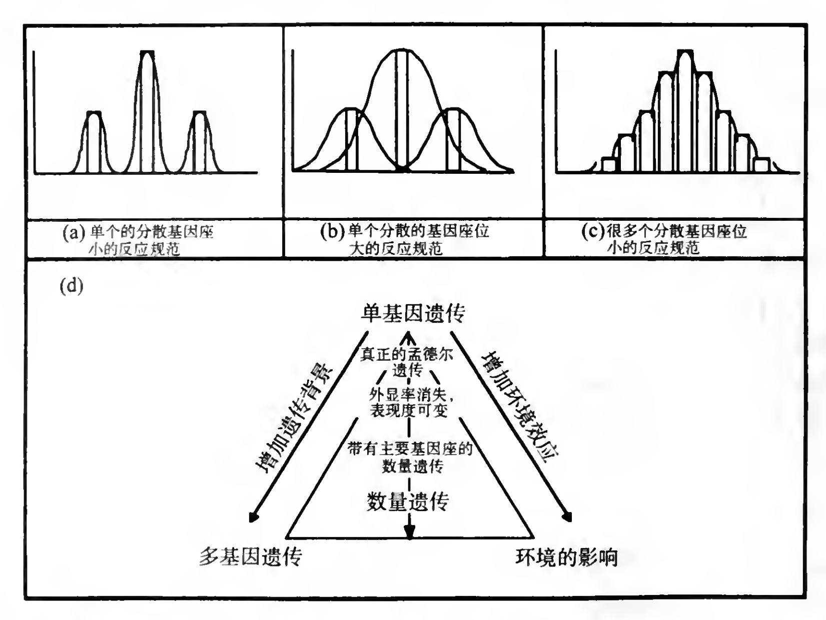
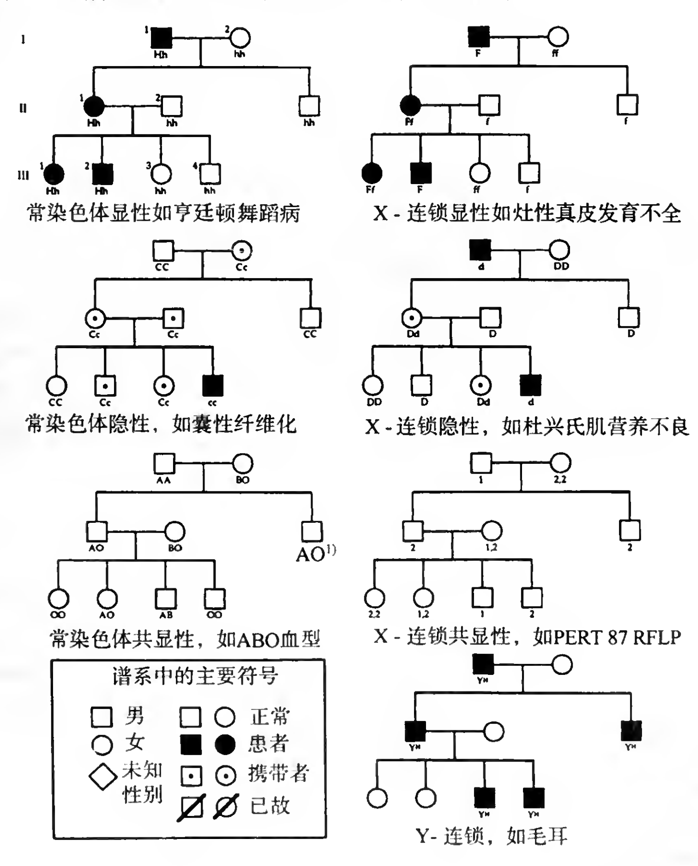


图1.9 环境对性状表型变异的影响。单个分散杂合座位产生3种基因型。(a)若反应规范很小,表型将成为不连续的性状。(b)若反应规范大,则变异是连续的。(c)若有多个分散座位涉及,则多基因模型预计基因型之间的差异将是小的;即使是小的反应规范各表型之间的差异是光滑的,结果产生连续变异。(d)少数性状是纯粹孟德尔式的、纯粹多基因或完全取决于环境的,大部分处在这两个极端之间,处在三角形的任何地方,基因数目和环境作用的增加使这些性状更不符合孟德尔式而更趋向数量遗传。

文框 1.1　孟德尔性状的家谱模式

人类性状的孟德尔谱系　在生物中如人类大规模的交配是不可能的。遗传的模型也不可能通过后代的比例来建立。取而代之的是利用家谱;而遗传的模型必须通过统计分析来估计(因为大部分人类的家庭是很小的,有时要建立一个清晰的遗传谱系是困难的,尤其是比较常染色体和性-连锁显性性状时。人类性状有一些基本的谱系:常染色体显性、隐性和共显性;X-连锁显性、隐性、共显性,以及Y-连锁。分居于X-和Y-染色体同源区的座位是以正常的常染色体方式遗传的,因为等位基因是由各亲本传递下来的——这就称为**拟常染色体遗传**(pseudoautosomal inheritance))。

基本遗传谱系的难题　即使一个特殊性状的遗传模型是清楚的,通过一个小的样本对谱系的解释也是困难的。通过一些因素降低了特定性状的表现度和外显率,从而进一步出现了难题,这些在大量的杂交中常能被忽略不计,但对于很多的谱系存在着严格的限制,因为单个病例可以使整个的谱系被曲解。这些难题中的某些完全反映了遗传机制(例如随机克隆失活的X-染色体、印迹出现、新的突变,X-连锁男性致死和胚系嵌镶),同时也可能由于遗传和环境因素两者的作用(即遗传背景、环境干扰和发育噪声)。探索谱系中最大的难题之一是**早现遗传**(anticipation)——某些性状在连续的几代中开始增加严重性和/或减少寿命。最近在一些人类疾病中对于早现遗传的分子基础已有阐述,反映了病原基因间三联体重复的行为(文框 15.2)



1)此处原文为“OO”型,可能有误,应为“AB”型或“AO”型。——译者注

遗传的类型	主要特点
常染色体显性	通过两性传递和表达 患者至少有一个亲本也是患者
常染色体隐性	通过两性传递和表达 患者的双亲通常都正常，但为**携带者**(carrier，无症状但带有隐性等位基因) 若双亲之间有血缘关系的话发生会增加
常染色体共显性	通过两性传递和表达 个体从双亲处各继承一个等位基因
X-连锁显性	通过两性传递和表达，但在女性中较多 男性患者把性状传递给所有的女儿 女性患者通常将性状传递给50%的儿子和女儿
X-连锁隐性	通过两性传递和表达，但一般都在男性中，由于男性是半合子 男性患者的双亲通常是正常的，但母亲是携带者，偶尔母亲也是患者(即纯合子)
X-连锁共显性	通过两性传递和表达 父亲的X-连锁等位基因总是传给女儿，不会传给儿子 母亲的等位基因可以通过女儿和儿子来遗传
Y-连锁	仅在男性中遗传和表达(限雄遗传) 患者的父亲和儿子都是患者

注意在家谱中代数是用罗马数字标出的，每代中的个体用其左侧的数字标明，这仅在第一个家谱中显示出。

文框1.2 遗传和环境方差的分解

表型方差的分解 表型方差(V_p)是在特定群体中对于一个特定的生物学性状来说所观察的总的方差。它可被分成两个主要的成分，**遗传方差**(genetic variance)(V_G)，由不同基因型所贡献的方差(即基因型之间的方差)。**环境方差**(environmental variance)(V_E)，反映所有外部的影响并对各种基因型产生反应规范(即基因型中的变异)。更进一步的成分是**基因和环境相互作用的方差**(gene-environment interactive variance)(V_{GE})，它反映当遗传方差和环境方差被计算出且从总的表型方差中被减去后余下的部分。V_{GE}可以视为两种成分之间相互作用的结果，但实际上直接计算是困难的而且常被忽略。用下列等式可表达这种关系

$$V_P = V_G + V_E + (V_{GE})$$

遗传方差和环境方差可被分解成一系列亚成分。

遗传方差 遗传方差(V_G)是表型方差的一部分，是由个体之间的基因型差异产生的。它可被进一步分成三种成分。

(1) V_A是**加性方差**(additive variance)(也叫做**基因方差**(genic variance))或者**育种值**(breeding value)。这反映对于特定性状具有加性效应的座位不同等位基因的作用。加性方差是用于选择育种的表型方差的基本成分。

(2) V_D是**显性方差**(dominance variance)，它反映了在每个座位上由等位基因相互作用而产生的效应。

(3) V_I是**互作方差**(interactive variance)，它反映了非等位基因互作产生的效应(即上位、抑制，增强作用)，而不是加性效应。

后面的两个成分通常一道聚集成"非加性方差",这是由于很难精确将它们分离出。对于一个特定性状的加性方差和非加性方差的相对量对动植物育种来说是特别有趣的,这种育种总是想选择人工选择中最有效的方法。将遗传方差分开这样可用下面公式来表达:

$$V_G = V_A + V_D + V_I$$

遗传背景(genetic background)　**遗传背景**是用来描述非特异性遗传效应的一个术语,这种效应改变了特定基因的表达。简单性状的遗传背景的效应导致了可变的表现度和外显率,包括非等位基因的相互作用(表 1.3)和位置效应,它常常影响汇集的转基因的表达以及涉及大量基因组重排的基因表达。在数量性状中任何由遗传背景所引起的变异都将由成分 V_I 来描述。

环境方差　像遗传方差一样,环境方差也能被分成为一些亚成分。

(1) $V_{E(g)}$**是一般的环境方差**(general environmental variance),反映特定群体中各部分都能显示出这种因素。

(2) $V_{E(s)}$**是特殊的环境方差**(special environmental variance),反映了仅在特殊的一组个体显示出的因素,如在动物怀孕时母体环境和共同的家庭环境。特殊的环境方差使家族和可遗传的性状难以区分(见下)。

进一步的成分可以被认为是环境的一部分或者作为方差的一种个别的来源。这种成分是**发育噪音**(developmental noise),它反映的是纯粹推测的事件。以分子水平来分,在不同的细胞中可能影响基因表达的一些因素。要区别发育噪音和环境引起的方差常常是困难的,但若一个性状能在身体的每一侧留下印迹的话。遗传方差和环境方差两者都不能留下这种印迹,而噪音却能全部被保留。环境方差的一部分可以用下面的公式来表达:

$$V_E = V_{E(g)} + V_{E(S)} + \text{发育噪声}$$

环境因素和噪声的效应就象遗传背景一样干扰简单性状的表现度和外显率。

表型模拟(phenocopy)　表型模拟是纯粹通过环境修饰而产生的情况。例如 *sonic hedgehog* 基因敲除小鼠表型少了头部和中部的结构。表型模拟可以通过使怀孕啮齿动物不供给胆固醇而产生。正常情况下,胆固醇可以与 Shh 蛋白结合使此蛋白具有功能。在此例中突变基因的作用可以通过从其环境中去除一种成分而得到模拟,这种环境对于基因产物的功能来说是必要的。

家族性和遗传力(familiality and heritability)　遗传力这个术语是用来表达表型方差中遗传额定贡献大小。在实验动物中,一个特定数量遗传性状的遗传力的论证是简单的。个体是群体缩小到极端,因此对于每个亚群体来说,一个性状的平均表型值离群体的平均值是很远的。每个亚群可相互杂交而后代被记录。如果一个性状是可以遗传的,那么后代的平均值和他们亲代的相似(即最初的群体),同时若观察初始群体的方差是在完全自然环境中,那么新群体的表型平均值是彼此相同,而且也和初始群体的相同。

不用繁殖而来确定遗传力的方法是寻找亲属间的相似性,亲属比群体中随机个体具有更多的机会获得相同的基因,而表型的协方差将反映相似遗传背景。然而亲属趋向于分享共同的环境,就象分获共同的基因一样。因此决定环境的贡献对于产生方差是否有意义是重要的。亲属们所分获的性状被描述为**家族性**的(familial),但不是所有家族性的性状都是遗传的,例如孩子像他们亲属那样说相同的语言,而语言在一个家庭中虽然是一种性状,但它并不遗传:一个生于英国家庭长在法国家庭的孩子将说法语。区别遗传力和家族性之间差别的方法是在大量不同的环境中观察表型性质。这对于一个可重复的性状来说是容易的(如绵羊的产毛量,奶牛的产奶量),但对于只表达一次的性状来说就比较困难(如谷类农作物的产量、人类的才智),这样论证人类数量性状

的遗传基础比较困难。但双生子领养的研究(将双生子放在不同的环境中生长来进行研究)将是有用的。

遗传力这个术语普遍用于两种含义。**广义遗传力**(heritability in the broad sense)(以 H^2 表示)也称为**遗传指数**(heritability index)(**H-统计的**,H-stastics)或**遗传决定的程度**(degree of genetic determination)。它表达为 $H^2 = V_G / V_P$[1)]。广义遗传力是计算在特定的环境中遗传方差和表型方差之比,而不计算基因对于特定性状发育的所有重要性。这样不会导致这个术语的大量误用,特别是用于人类的社会问题。**狭义遗传力**(heritability in the narrow sense)(用符号 h^2 表示)表达为 $h^2 = V_A / V_P$,是计算加性遗传方差对总的表型方差的比例。这样来判断所观察到的表型方差受选择育种影响的程度。人工选择是在一定环境一定的群体中进行的,从而增进商业价值的性状,这样就限制了广义遗传力的价值,使广义遗传力不能延伸到所有群体中,它对于育种来说是不重要的。

(徐晋麟 译)

1) 原文中这类公式都用破折号来连接等式的两侧,为明确起见,将破折号改为等号。——译者注

第2章
细 胞 周 期

基本概念和定义

- **细胞周期**(cell cycle)是连续的细胞分裂过程中发生事件的顺序。
- 在细胞周期中,许多过程协同作用,有些连续发生(如细胞生长),有些是非连续发生的事件和标志(如细胞分裂)。细胞分裂必须与细胞生长和 DNA 复制相谐调,这样才能保证细胞的大小和 DNA 含量保持恒定。
- 细胞周期由**核周期**或称为**染色体周期**(DNA 复制和分配)和**细胞质周期**或**细胞分裂周期**(细胞质成分的合成和分裂,在真核细胞中还包括细胞器)组成。每个营养期的细胞仅有一份或两份 DNA 拷贝,DNA 被认为是从其他细胞中得到的,DNA 的复制和分配必须得到精确的控制。在母代细胞大得应该分裂时,细胞内的成分继续合成达到足够的量时,均等地分配到子代细胞中。**中心体**(centrosome)是个例外,中心体在染色体分离中起关键作用的细胞器,在有丝分裂前复制,和染色体一起分配到子代细胞中(**中心体周期**, centrosome cycle)。
- 在真核细胞中,复制和有丝分裂是染色体周期中的主要事件,它们因受到调节而不能自发进行。相反在细菌中的类似过程:复制和分配,在快速生长时受到谐调,使部分已复制的染色体得以分配到子代细胞中。真核细胞的细胞周期被严格分为不同时期,按一定顺序进行,而细菌的细胞周期的不同时期是可以相互重迭的。
- 真核细胞的细胞周期受关卡(checkpoint)控制,关卡的调控蛋白接收细胞周期本身的监视器(内在信息)和环境(外源信息)监视器的输入。内在信息是指细胞周期是否按照正确的顺序进行,即细胞周期中下一个时期是否在上一时期结束后开始。外在信息是指在外界环境不合适的情况下,细胞周期必须得到协调,及时使细胞周期停止。
- 蛋白激酶调控细胞周期。细胞周期的转换涉及到激酶活性突然爆发的正反馈调节,使一系列效应蛋白磷酸化。如果内部或外部环境不适合,这些激酶活性在细胞周期关卡处被抑制,DNA 复制和有丝分裂的转换是负反馈调控,未完成 DNA 复制抑制有丝分裂,在有丝分裂进行的时候,通过磷酸化和灭活复制所需蛋白,阻断 DNA 复制的起始。细胞周期是一个复杂的信息网络的结果,在信息网络中,激酶的活性受多项正负信号的综合控制。

2.1 细菌的细胞周期

DNA 复制和协同生长　细菌的染色体周期根据 **Helmstetter-Cooper 模型**(或称为 **I+C+D 模型**)分为**间期**(interval phase)、**染色体复制期**(chromosome replication phase)和**分裂期**(division phase),分别用字母 I、C 和 D 表示。

DNA 复制发生在 C 期,时间固定(大肠杆菌中约为 40 分钟),反映了复制整个染色体所需的时间(参见 26 章,复制)。DNA 复制结束后 D 期就开始了,并在细胞分裂时达到高峰,D 期时间也固定(在大肠杆菌中约为 20 分钟),被认为是合成细胞分裂所需细胞成分的时间,大肠杆菌染色体周期的最短时间为 1 小时。

因为 C 期和 D 期的时间相对固定,细胞分裂的时间变化反映了连续复制起始之间 I 期的时间变化。大肠杆菌的倍增时间长至 3 小时,短至 20 分钟。在生长缓慢的时候,I>C+D,复制在细胞分裂之前结束。在生长旺盛时,倍增时间比一轮完整的复制和分裂的时间要短。细胞只有使 I<C+D,即新一轮的复制必须在上一轮复制完成前开始,才能使固定的染色体周期适应快速生长的需要。此外,在快速生长中,子代细胞得到的染色体已部分复制[**多叉染色体**(multiforked chromosomes)],所以复制能在下一轮细胞分裂前完成。

复制起始的频率被正调控因子所调节,这种调控因子在每个*复制起始*(参阅)处,必须达到一定的浓度才能使复制起始。在快速生长过程中,调节因子积累更快,允许更频繁的起始。一旦复制起始,在分裂中细胞的复制起始点和调控因子的有效浓度就失效了,所以在下一轮起始时,这些因素必须重新积累。支持存在正调控因子的根据是复制起始是依赖于蛋白质的从新合成的,但到目前为止,这正调控因子的天然属性还不得而知。复制起始蛋白 DnaA 可能是这种调控因子的候选者,另外控制 *oriC*(和 *dnaA* 启动子)甲基化的因子可能也与此相关(参见*复制起始*、*甲基化*)。

分配和胞质分裂　染色体周期达到高潮的标志是已复制的染色体进行分配,接着就开始细胞分裂。在母细胞的中间形成被称为**周隔环**(periseptal ánnulus)的**隔片**(septum),在此处细胞周缘的内外膜相互融合。通过复制形成的环迁移到细胞长度的四分之一和四分之三位置,这是下一轮细胞分裂子代细胞分离的地方。一旦隔片形成,细胞就进行**胞质分裂**(cytokinesis)——以二分裂(binary fission)进行分裂。

研究有关阻断细胞分裂和分配突变表明,有两个过程可以被阻断,这样的突变可分为几类。Fts 突变体隔片形成有障碍,丝体形成是温度敏感性的(*f*ilament that are *t*emperature *s*ensitive, *fts*)。丝体上往往有规则排列的类核,这说明复制和分配功能是正常的。*min* 突变是隔片形成太频繁,结果形成很多**小细胞**(minicells),有些小细胞不含有染色体 DNA(有可能含有质粒 DNA)。*par* 突变体细胞大小正常,但在染色体分配有缺陷,且高频率地产生二倍体和无核细胞的频率增加。

细胞分裂和分配调节途径的细节还有待于进一步研究,但目前已知其中一些关键分子。细胞分裂的起始因子可能是 FtsZ,FtsZ 从结构和功能上都与在真核细胞中形成收缩环的微管蛋白相似。FtsZ 在整个细胞周期中普遍存在,分布在 D 期环周围形成 **Z 环**(Z ring)。FtsZ 的含量与细胞分裂的频率相一致,*fts*Z 突变就不能形成隔片(和丝体),而过

度表达会使隔片产生过多,形成小细胞。ZipA 蛋白可能对 FtsZ 的定位很重要,它的 N 端定位在细胞膜上,C 端与 FtsZ 相互作用。到目前为止隔片在细胞中是如何定位的还不清楚。当温度敏感的 *fts*Z 突变体转移到适宜的温度时,丝体可以迅速形成收缩环,所以可能存在**成核位置**(nucleation site)。基因 *minB* 可以限制染色体分配到中心环,抑制在前一轮细胞分裂后末端环形成的过程。MinC 和 MinD 是分配抑制因子,而 MinE 是 MinC 和 MinD 的拮抗剂。这三种成分的平衡抑制末端分配,保护中心环免受抑制。

细菌染色体分配过程与质粒分配(参阅)类似。分配涉及染色体与细胞膜的相连,并有可能被复制所调节:复制的起始和终止,复制叉的激活都是与细胞膜相关的过程。细菌中影响分配的突变体可分为两类:一类与互联的已复制的染色体作用(包括拓扑异构酶和 Xer 位点特异重组突变体);另一类直接影响分配过程本身。在后一种情况中,在染色体上未发现顺式作用位点,但有一些反式作用因子被发现,如膜蛋白 MukA 和微管相关蛋白 MukB,这两个基因突变会形成无核细胞。

2.2 真核细胞的细胞周期

标准的真核细胞的细胞周期　　真核细胞的细胞周期可以分为四个互不重叠的阶段。染色体周期中独立的事件(DNA 合成和有丝分裂)分别发生在 **S 期**和 **M 期**,在大多数细胞周期中,M 期和 S 期为 **G_1 和 G_2 期**(间期 1 和 2,也分别称为合成前期和合成后期)所分开,在 G_1 和 G_2 期中 mRNA 和蛋白质继续积累(图 2.1)。从细胞周期中的一个期进入另一期称为**细胞周期过渡**(cell cycle transition)。整个细胞周期中,在有丝分裂期中可以观察到明显的细胞结构的变化,而其他时期肉眼看来毫无变化,称为分裂**间期**(interphase)。

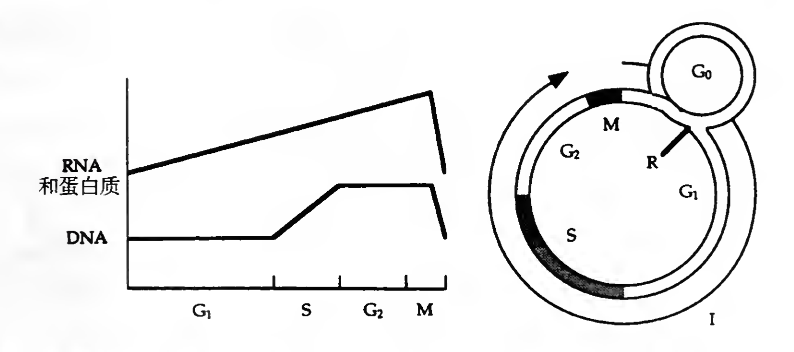


图 2.1　标准的真核生物细胞细胞周期。染色体周期分为四个阶段:构成**间期**(interphase)的 G_1, S 和 G_2 期;有丝分裂的 M 期。左边是细胞周期不同阶段的相对时间,根据不同物种、细胞类型和生长情况而不同。如果在 G_1 期早期生长因子缺乏,动物细胞可以进入静止期,G_0;如果细胞已通过限制点(R),就进入下一轮 DNA 复制和分裂。图中反映了细胞成分的持续积累和 DNA 非持续合成。当 M 期结束细胞分裂完成,细胞成分减半。

细胞周期中的变化(表 2.1)包括:一个或两个间期(gap)缺失,S 期或 M 期的缺少会分别导致细胞中 DNA 含量的减少或加倍。此外,细胞可以在细胞周期中的某一阶段**被阻滞**(arrested),如在卵母细胞的成熟过程中和有丝分裂期后的(postmitotic)细胞(如神经元细胞)。动物细胞还可以完全从细胞周期中离开,进入一个生长和分裂受抑制的静止期,称为 G_0 期。这些反映了必需存在相应的生长因子和环境中的信号分子以及其他水平

表 2.1 真核细胞细胞周期四个阶段的变化

改变	环境
阶段被去除	
没有 G 期	因为在卵细胞中有足够的供细胞分裂所需的原料，所以在一些有大型卵细胞的动物发育过程中没有 G_1 期，M 期和 S 期得以快速转换。
	许多组织缺少一个或整个 G 期，如网柄菌属盘脉菌（*Dictyostelium discoideum*）在有丝分裂后立刻进行 DNA 的复制（没有 G_1 期）；在酿酒酵母中，DNA 复制后立刻进行有丝分裂（没有 G_2 期）。
没有 S 期	在减数分裂时，两轮细胞分裂中间没有 DNA 的合成
没有 M 期	在果蝇（*Drosophila*）分泌组织中有多轮 DNA 复制但细胞并不分裂，以产生多线染色体（polytene chromosomes）（参阅）。
阶段延长	
G_1，G_2 或 M 期阻滞	卵母细胞根据物种不同，在 G_1，G_2 或 M 期阻滞，受精可以脱离阻滞，使细胞周期继续进行。
细胞周期从 G_1 期中退出（G_0 期）	很多动物的细胞可以从 G_1 期中退出，进入静止期，常称为 G_0 期，G_0 期可数月或数年。这往往发生在 G_1 期的早期关键的生长因子缺乏或细胞周期调控机制失控时。如果所需的生长因子具备，静止的细胞可以重新进入细胞周期。因为调节成分需要重新合成，所以在 S 期开始前需要有较长时间。正常的体细胞经常经过多轮分裂后进入 G_0 期，称为 **Hayflick 限制**（Hayflick limit）。又称为**细胞衰老**（senescence），G_0 在很多动物中可能与端粒的长度有关，在真菌中也可被某些质粒诱导。有些细胞（如神经元和肌细胞）作为它们分化的一部分，完全从细胞周期中脱离，成为**有丝分裂期后的**（postmitotic）细胞。

的对细胞周期的调节，才能使单个细胞的分裂与多细胞组织相协调。这些调节机制的失控可以导致肿瘤中异常细胞分裂（参见癌基因和癌一章）。

细胞周期关卡 细胞周期基本功能就是精确地复制基因组、均等地分配到两个子细胞中。正是这个原因，细胞周期的各个事件按正确的顺序进行就极为重要，也就是说细胞周期中下一期必需在上一期完成后才开始。细胞内 DNA 含量保持恒定，只有在 DNA 复制和有丝分裂交替进行，有丝分裂在 DNA 复制完成后开始，复制在有丝分裂精确地将 DNA 分配到子代细胞中后开始。细胞通过将细胞周期组织成一系列相互依赖的事件来满足这一要求。如果有丝分裂被阻断，细胞周期就停留在 M 期直到阻断因素去除——细胞周期不会自行进行下去并复制 DNA（DNA 的复制依赖于有丝分裂的完成）。同样当 DNA 复制受阻，细胞也不会试图进行有丝分裂，因为有丝分裂依赖于 DNA 复制的完成。

细胞周期进一步的功能是使染色体周期与细胞生长相协调，这样细胞质就不会增加或减少，并且细胞也不会在不适合的环境下继续生长。细胞周期的进行还依赖于细胞的大小，并为所提供的营养、交配信息素（在酵母中）、生长因子和激素（动物中）所调节。

细胞拥有一系列的调节系统，来感受细胞周期进行的情况，当出现失常时，在相应阶段终止细胞周期。这些调节机制称为细胞周期的**关卡**（cell cycle checkpoint），代表调节细胞周期的内在信号系统。关卡机制也可对外源信号作出反应，如当营养或生长因子缺乏时，时细胞周期受阻。在细胞周期中有很多关卡，被分为两大类——发生在 G_1 期，对进

入S期进行调节；发生在 G_2 期，调节细胞进入M期(图2.2)。这两类阻断暗示内源和外源信号可能集中到细胞周期调节的共同成分中。另一类关卡保证组成有丝分裂的事件能有序且相互依赖地进行。

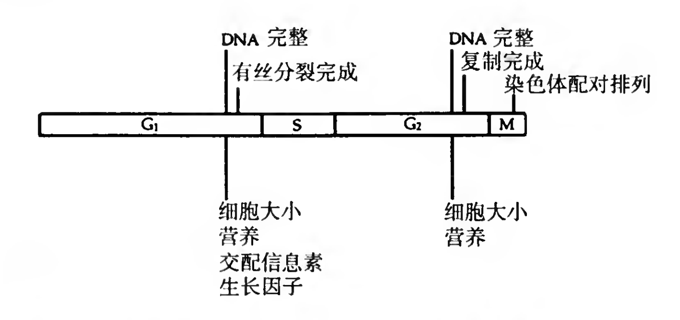


图2.2 已知真核细胞周期的关卡。代表了特异的蛋白激酶的激活/灭活的位置。

G_1 和 G_2 期的关卡对不同器官有不同程度的意义，这反映了细胞从环境中得到的信息。在啤酒酵母中 G_1 期的关卡占主导地位，称为**起始**(START)，在动物细胞中则称为**限制点**(restriction point)或**关键点**(commitment point)。在 G_1 期，酵母对营养和交配信息素的提供情况进行判断，动物细胞对生长因子作出反应。如果环境不利，细胞就停留在关卡处，一旦通过了关卡，新一轮的DNA复制和有丝分裂得以进行，并不再考虑环境的因素。相反，在裂殖粟酒酵母(*Schizosaccharomyces pombe*)中，环境的因素在 G_2 期的关卡进行检测。在适合的条件下，细胞进行有丝分裂和下一轮的DNA复制直到遇到关卡为止。对单倍体酵母细胞而言，停留在 G_2 期比 G_1 期的优点是在 G_2 期基因组有两份拷贝，DNA损伤可以通过重组来修复。

细胞周期调节的研究　两种互补的研究手段被用来分析和提取细胞周期的调节成分。在异核体的方法中，不同染色体周期的细胞核与正常细胞质融合，观察其行为。培养的哺乳动物细胞和两栖类动物的卵细胞常被用来做这类实验。表2.2表示的是不同细胞周期阶段的同步化的成纤维细胞的结果。**M期激活因子**(M-phase promoting factor)存在的最早证据是M期细胞可以诱导任何间期细胞进入有丝分裂。在非洲爪蟾细胞核移植实验中也可以得到类似的结论——当植入一处于M期卵细胞时，间期细胞形成纺锤体，从这些细胞中得到的细胞质还可以诱导 G_2 期的卵母细胞进入M期。从卵细胞中提取得到的相关因子，被称为**成熟促进因子**(maturation promoting factor, MPF)。进一步的研究表明，成熟促进因子也可以诱导体细胞的有丝分裂，也就是M期激活因子，统称为**MPF**。

另一种方法是利用酵母遗传的多变性，分离与细胞周期功能相关的条件突变体。很多不同细胞周期阶段阻断的突变体被发现，统称为 ***cdc* 突变体**(细胞分裂周期)，同时这样的突变体的数量还在不断地增加。另一类称为 ***wee* 突变体**，过早通过细胞大小的关卡，所以细胞比野生性细胞要小。许多从 *cdc* 突变中得到的基于并非是细胞周期专一性的调节因子，而是细胞周期所依赖的过程，如DNA复制、修复和配对中的调控因子。尽管如此，还是有一些 *cdc* 基因是直接调控细胞周期的。令人欣慰的是以上两种方法得到的MPF因子基本是同一系列的分子。

表 2.2　异核体实验研究调控细胞周期因子

融合	结果	结论
S+G_1	两个核都复制	S期的核含有启动进入S期的激活因子
S+G_2	S期细胞完成复制，G_2 期的细胞等待S期核完成复制，两个细胞都进入M期	G_2 期对S期启动因子没有反应，同时S期启动因子也是M期的抑制因素
M+G_1，S或G_2	间期细胞不管染色体复制情况，进入有丝分裂	M期细胞核含有进入M期的激活因子
G_1+G_2	细胞并不进入有丝分裂或复制	S期和M期激活因子的表达是瞬时的

2.3　细胞周期调节的分子基础

细胞周期蛋白和细胞周期蛋白依赖激酶　　细胞周期阶段性进行反映了调节不同细胞周期事件中的关键分子的不同磷酸化状态。这些不同的磷酸化状态代表了细胞周期的过渡。G_1－S的过渡涉及DNA复制相关蛋白的磷酸化，而G_2－M期的过渡与有丝分裂相关蛋白的磷酸化。细胞周期调控的基础是一系列激酶家族，将相应细胞周期过渡所需蛋白磷酸化。

发现细胞周期调节涉及蛋白激酶是在酿酒酵母 *cdc* 突变体的研究中，发现阻断起点(START)的基因产物 *CDC28*，是一 34kD 的蛋白激酶，是G_1－S过渡的主要调节因子。在粟酒酵母(*S. pombe*)中同样重要的G_2－M的调节因子是 *cdc2*，是上述激酶的同源蛋白。在这之后，脊椎动物细胞中编码类似蛋白激酶的基因逐步被克隆，这些蛋白都可以恢复酵母 *cdc* 突变体的细胞周期功能。值得一提的是，非洲爪蟾的CDC28/Cdc2的同源蛋白就是MPF的成分。

这些激酶在细胞核中含量保持恒定，但随细胞周期的进行而活性发生变化。对它们活性变化的解释来自于对海胆卵的研究，从中发现了一类分子，它们的合成和活性随细胞周期而变动。这些分子被称为**细胞周期蛋白**(cyclin)，此后在其他包括酵母和脊椎动物的真核生物中也发现了cyclin。MPF的第二类成分是B型cyclin。

MPF的活性在有催化作用的激酶亚基，但依赖于细胞周期蛋白亚基的构象改变而促进其活性。所以细胞周期激酶被称为**细胞周期蛋白依赖激酶**(cyclin-dependent kinase, CDKs)，以CDK-cyclin全酶形式发挥作用。这种细胞周期调节方式在真核细胞中是保守的。

酵母和动物细胞周期中 CDK-cyclin 的差异　　从酵母中得到很多潜在的CDK，但只有酿酒酵母的CDC28和粟酒酵母的Cdc2看来是直接参与细胞周期调控，分别为G_1－S和G_2－M的过渡所需。动物细胞中有更多的CDKs。第一个被发现的，可能是MPF的分子是$p34^{CDC28/Cdc2}$，专一性在G_2－M的过渡中发挥作用。动物细胞中发现有十多个CDKs，其中五个在细胞周期早期阶段发挥作用。

因为在动物和酵母中,不同的细胞周期阶段合成不同的细胞周期蛋白,所以细胞周期蛋白的多样性更甚于 CDKs。至少有八个家族的脊椎动物的细胞周期蛋白(命名为 A - H)。因为 CDKs 在细胞周期的不同阶段过渡时作用于不同的底物,所以细胞周期蛋白不仅为激酶活性所需,还为底物专一性所需。在动物中,不同类型的细胞表达不同的细胞周期蛋白,这样有助于不同分化细胞对细胞周期特异的调控,在所有器官中主要有三种类型的细胞周期蛋白:**G_1 期细胞周期蛋白**调节 G_1 - S 的过渡;**S 期细胞周期蛋白**为 DNA 复制所需;**M 期细胞周期蛋白**在有丝分裂中发挥作用。M 期细胞周期蛋白包括酿酒酵母(*S. cerevisiae*)CLB 细胞周期蛋白、脊椎动物 A 型和 B 型细胞周期蛋白以及粟酒酵母的 Cig13。它们是稳定的蛋白,都有一个称为**破坏盒**(destruction box)的保守基序,破坏盒是触发泛肽化(参阅)所需,使蛋白在有丝分裂过程中降解。其他细胞周期蛋白在遗传上不稳定,含有 PEST 结构域(参阅),它们的表达水平由其基因的转录激活所决定。所有的细胞周期蛋白都有保守的 CDK 结合所需的**细胞周期蛋白盒**(cyclin box),参见图 2.3。酵母和动物细胞的 CDKs 和细胞周期蛋白在表 2.3 中得到总结。

表 2.3 酵母和哺乳动物不同细胞周期阶段的 CDKs 和细胞周期蛋白

阶段	酿酒酵母	粟酒酵母	哺乳动物
G_1	CDK:CDC28 Cyclin:CLN1 - 3	CDK:Cdc2 Cyclins:Cig2	CDK:CDK2,4,5,6 Cyclins:D1 - 3 CDK:CDK2 Cyclins:E 型
S 期	CDK:CDC28 Cyclins:CLN5,CLN6	CDK:Cdc2 Cyclins:Cig2 ?	CDK:CDK2 Cyclins:A 型
G_2/M 期	CDK:CDC28 Cyclins:CLB1 - 4	CDK:Cdc2 Cyclins:Cdc13	CDK:CDK1(CDC2) Cyclins:A 和 B 型

在哺乳动物细胞周期的 G_1 期中的两种 CDK -周期蛋白系统是有活性的。CDK2 -周期蛋白 E 复合体对于 G_1 - S 过渡是必需的。其他的 CDKs 和周期蛋白 D 负责解释来自环境的生长因子信号,作用限制位点开辟了细胞进入新的 G_1 期和 G_0 期。哺乳动物的 CDK -细胞周期 B 复合体是 MPF -脊椎动物和酵母 Cdc2/CDC28 的同源物,称为 Cdc2 或 CDK1。CDK7/H -周期蛋白是 CAK,哺乳动物 CDK Thr - 161 激酶,CDK7/H -周期蛋白被认为是连续表达的。

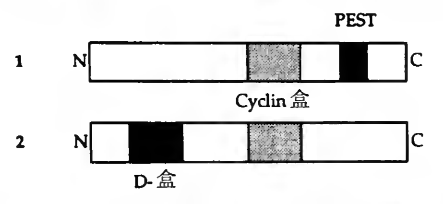


图 2.3 cyclin 的结构域。(1)含 PEST 结构域的 cyclin 易被蛋白降解,很不稳定。这一类包括酿酒酵母的 CLN cyclin 和脊椎动物的 C、D 和 F 型 cyclin。大部分是 G_1/S cyclin。(2)有丝分裂 cyclin 在间期中比较稳定,包含一 M 期泛肽依赖降解所需的破坏盒。第一个 cyclin 是根据其活性的变化分离得到的,但有些在体内表达稳定,根据 cyclin 同源性定义。

CDK-cyclin 活性的调节 细胞周期转换最显著的特点是 CDK 活性的爆发,它骤然打开细胞周期事件中靶蛋白进入磷酸化状态的开关。因为细胞周期蛋白在细胞中是逐步积

累，只有有丝分裂中细胞周期蛋白快速被蛋白酶降解，所有激酶活性的爆发并不是通过细胞周期蛋白的合成和降解来调节的。

CDKs 磷酸化靶蛋白，但其本身也为磷酸化所调控(参见第 28 章信号转导)。酵母的 CDC28 和 Cdc2 的 CDKs 在两个关键的残基 Tyr - 15 和 Thr - 161 上磷酸化，。Thr - 161 的磷酸化是激酶活性所需，但 Tyr - 15 的磷酸化可以抑制 Thr - 161 的磷酸化。在酵母中决定 CDK-cyclin 活性的是 Tyr - 15 的磷酸化状态，一些上游调控成分已被发现。在粟酒酵母中，Wee1 是可以磷酸化 Cdc2 Tyr - 15 的酪氨酸激酶，并灭活 Cdc2。Wee1 活性为 Cdc25 磷酸脂酶所拮抗，Cdc25 可以将同样底物上的磷酸基团移去。Wee1 和 Cdc25 本身为内外信号所调节，被认为是此物种中 G_2 - M 关卡的基础分子。是进入有丝分裂还是阻滞在 G_2 期反映了这两种活性的相对水平，对这一关卡的反应调节网络在下文中进行讨论。

在哺乳动物细胞中也发现了 *wee*1 和 *cdc 25* 的同源蛋白，在哺乳动物中情况更复杂，因为有针对不同 CDK-cyclin 复合物的多种异构体。哺乳动物的 CDC2(CDK1)有三个磷酸化位点，Thr - 14 和 Tyr - 15 的磷酸化抑制激酶活性，而 Thr - 161 的磷酸化是激酶活性所必需的。使 Thr - 161 磷酸化的蛋白已经被发现，是另一 CDK-cyclin 复合物，由 CDK7 和 cyclin H 组成，称为 **CDK 激活激酶**(CDK-activating kinase，**CAK**)(参见TFIIH)。使 Thr - 14 位磷酸化的蛋白激酶目前还未发现。

CDK - cyclin 复合物也为抑制蛋白所调节。内源和外源的调节通过三个水平来实行。如下文将讨论的，粟酒酵母 Rum1 蛋白是一专一性有丝分裂 CDK-cyclin 复合物的抑制蛋白，在整个 G_1 和 S 期合成，防止细胞周期越过 DNA 复制进入有丝分裂。酿酒酵母的 FAR1 蛋白被交配型信息素激活，抑制起点 CDK-cyclin 复合物，使细胞处于 G_1 期，等待交配的进行。哺乳动物中发现两个家族的 **CDK-cyclin 抑制剂**(CDK-cyclin inhibitors，**CKI**)。一类抑制所有的 CDK-cyclin 活性，另一类专一性抑制含 CDK4 和 CDK6 的 D 型细胞周期蛋白复合物(它们的主要底物是成视网膜细胞瘤易感蛋白，Rb)。这些调节机制将在下文中详细讨论。

2.4 细胞周期的进行

酵母中起点关卡的通过　　起点(START)是酿酒酵母中主要的细胞周期关卡，在此物种已被很好地阐明(图 2.4)。在 G_1 期，CDC28 的 CDK 可以和三种细胞周期蛋白(CLN1，CLN2，CLN3)中的任何一种结合。这三种 cyclin 至少在实验室条件下具有功能冗余性(三种中的一个或一对缺失后不影响表型，但三个同时突变使细胞阻滞在 G_1 期)。高度不稳定的 CLN3 的表达是组成型的，可以作为细胞生长的指示。当 CDC28-CLN3 浓度达到一定水平，激酶激活两个转录因子 SBF 和 MBF 的表达，这两个转录因子都有一共同的成分：SWI6。SBF 启动 G_1 cyclin 所需基因 CLN1 和 CLN2 的转录。这些细胞周期蛋白和 CDC28 结合形成 CDK-cyclin 复合物激活六个 CLB 细胞周期蛋白(有丝分裂细胞周期蛋白 CLB1，CLB2，CLB3，CLB4 和 S 期细胞周期蛋白 CLB5，CLB6)，这些细胞周期蛋白都有保守的破坏盒。首先它们减缓了 CLB 细胞周期蛋白的蛋白酶降解，其次增加了 CDK-CLB 抑制剂 SIC1 的水解速度。

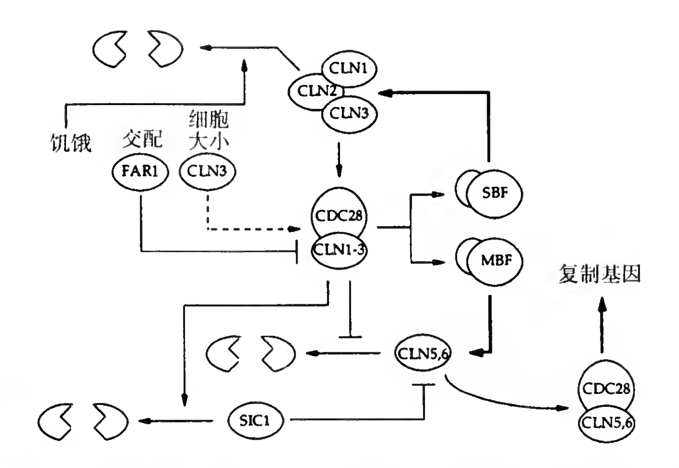


图 2.4 酿酒酵母中起点分子调节的总结。CLN3 是细胞大小的指标(点线表示),达到临界浓度后,CDC28 的激酶活性被诱导并激活两个转录因子。SBF 促进 *CLN-3* 转录,产生一正反馈通路。由于饥饿 CLN2 的蛋白酶降解能延缓进入 S 期。MBF 激活 CLN5 和 CLN6 基因,同时被激活的 CDC28 激酶抑制 cyclin 的降解并促进抑制剂 SIC1 的降解。CLN5 和 CLN6 与激活复制基因的表达的 CDC28 形成复合物。

第二个转录因子 MBF 启动 CLB5 和 CLB6 基因转录,通过刺激 S 期细胞周期蛋白的转录增加它们的稳定性。G_1 期专一性的 CDK-cyclin 复合物当细胞生长足够时开始进入 S 期。S 期专一性的 CDK-cyclin 复合物刺激复制蛋白的转录或激活,这可能包括建立识别酵母 ARS 元件上复合体的结构区(参见第 26 章复制)。

细胞只有在有丝分裂结束后才能通过 START 这一起点,在粟酒酵母中是通过一负反馈机制来完成的:有丝分裂 CDK-cyclin 复合物抑制 DNA 复制起始所需蛋白 Cdc18。在没有 Cdc18 活性时,DNA 的复制中断,当有丝分裂激酶被破坏后,抑制效应去除(见下文)。

哺乳动物细胞进入 S 期 在哺乳动物细胞中,四个 CDK(CDK2,CDK3,CDK4 和 CDK6)在 G_1 早期起作用。四个都与 D cyclin 结合,D cyclin 以生长因子依赖的形式(生长因子对细胞周期的调节在下文中讨论)在 G_1 期早期合成。CDK4 和 CDK6 在调节下游事件中起主要作用。目前所知的这些早期 G_1 期 CDK-cyclin 复合物的底物是成视网膜细胞瘤易感蛋白 RB－1,RB－1 是细胞周期负调控因子。非磷酸化状态的和转录因子 E2F 结合,E2F 正常情况下激活进入 S 期所需基因。RB－1 通过两种途径抑制这些基因的表达:阻断 E2F 的激活结构域和直接抑制染色质的重建。CDK－D cyclin 复合物对 RB－1 的磷酸化,去除了 RB－1 的抑制作用。CDK-D-cyclin 激活的一重要下游作用是 E cyclin 的激活,E cyclin 是 G_1－S 过渡所需的蛋白。E cyclin 和 CDK2 形成复合物,然后激活 CDC25A 磷酸酯酶,CDC25A 接着可以激活 S 期专一性的 CDK2/A cyclin 复合物。这一复合物是复制起始所需,可以与 PCNA(参阅)——真核细胞 DNA 聚合酶的成分之一——一起免疫共沉淀,定位于复制起始点。起始复合物的靶蛋白还有待进一步研究,cyclin A 和 cyclin E 复合物可以磷酸化和调节转录因子(如 E2F 家族、p53、B-Myb 和蛋白螺旋－环－螺旋抑制蛋白 Id2)的活性。

每一细胞周期只复制一次的限制 上面谈到的细胞融合表明 S 期的核可以诱导 G_1 期

核进入复制，但 G_2 或 M 期的核不行，只有完成有丝分裂后才有 DNA 复制的能力。不管在何种情况，复制在一个细胞周期中只进行一次。一个试图解释这些结果的模型涉及**复制许可因子**(replication licensing factor)，有如下特性：

- 在有丝分裂后期和 DNA 复制起始点相互作用
- 为 DNA 复制起始所必需，但一旦复制起始它就被灭活或降解
- 不能通过核膜

这种模型假设复制许可因子在有丝分裂末期，核膜重组之前与 DNA 复制起始点结合。这种因子为刺激进入 S 期的 CDK-cyclin 复合物的调节目标，在 S 期开始时启动复制起始。在细胞周期的其他时间内，新的许可因子不能进入细胞核，重复的复制起始被抑制。因子在有丝分裂开始时进入细胞核，保持非活性状态或因为染色质凝聚状态的原因不能与 DNA 相互作用。当染色质构象去凝聚状态时，因子与复制起始点作用，但仍保持非活性状态，等待 S 期信号。核膜形成阻止其他因子进入细胞核，任何核内未结合的和细胞质中的因子就被降解。

在非洲爪蟾属(*Xenopus*)、哺乳动物和酵母中，很多有许可因子特性的蛋白已被发现。在酿酒酵母中，*CDC46* 基因产物就被认为是许可因子或其中的一个成分。尽管酵母的细胞核膜在有丝分裂的过程中并不破坏而许可因子通过转运可进入细胞核。

控制进入有丝分裂　在酵母中调节起点 START 的细胞周期蛋白依赖激酶(Cdc2, CDC28)也同样控制进入有丝分裂，但与有丝分裂的细胞周期蛋白结合。有丝分裂的细胞周期蛋白在酿酒酵母中由 *CLB1*，*CLB2*，*CLB3* 和 *CLB4* 基因编码，在粟酒酵母中由 *cdc13* 基因编码。在脊椎动物中，CDK1(CDC2)激酶与酵母中相对应的 A 和 B 型细胞周期蛋白都有有丝分裂破坏盒。在粟酒酵母中发现了很多与调节 G_2 - M 期关卡相关蛋白，在真核细胞中也得到了类似蛋白。最关键的是，有丝分裂通过翻译后正反馈机制，使 Cdc2 活性爆发启动(这与在酿酒酵母中通过转录后反馈机制进入 S 期不同)。这种机制在整个真核生物中是保守的，但进入有丝分裂关卡的调节如下文讨论的是通过不同途径。

在粟酒酵母，Cdc2 细胞周期依赖激酶主要为 Tyr - 15 的磷酸化调节(图 2.5)。它由 Cdc25 和另一个磷酸酯酶 Pyp3 去磷酸化激活，而被激酶 Wee1 和 Mik1 灭活。G_2 期到 M 期的转换与 Cdc2 激酶活性的破坏相一致，Cdc2 激酶由一个正反馈途径来调节，在此途径中，Cdc2/Cdc13 激酶磷酸化并激活 Cdc25，再使 Cdc2 去磷酸化和激活(Cdc2 也可使 Wee1 失活)。这样激酶活性的突然增加促进了很多靶蛋白的磷酸化状态的改变，有效地转换了启动有丝分裂的开关(表 2.4)。Wee1 和 Cdc25 加工内源信号并汇集了来自环境的信号，可能是主要的关卡调节因子。激酶 Nim1 对细胞的营养状态作出反应，调节 Wee1 的活性。一组基因，包括 *cdc18*，*cdt1* 和 *cut5*，监测 DNA 复制是否完成和有无受损 DNA 的存在，它们的蛋白产物被认为形成一识别复合物，为 Rad1，Rad3，Rad8，Rad17，Rad26 和 Hus1 检测。这些信号的下游目标可能是 Cdc25 或 Wee1，或者两者都是。最近发现 Rad3 磷酸化 Chk1，Chk1 进一步磷酸化 Cdc25，使其不能与 14 - 3 - 3 蛋白形成复合物。在哺乳动物细胞中，还有不依赖于 Tyr - 15 磷酸化的调节 CDC2 激酶活性的调节机制。

酿酒酵母缺乏明确的 G_2 期，DNA 损伤和 DNA 复制的关卡与上述不同。在这种物种中，细胞形成纺锤体，对 DNA 损伤和未完成的 DNA 复制的反应是在中期阻滞。关卡调控的靶是后期促进复合物(APC，见下文)，它是通过泛素与有丝分裂的 cyclin 破坏盒相连

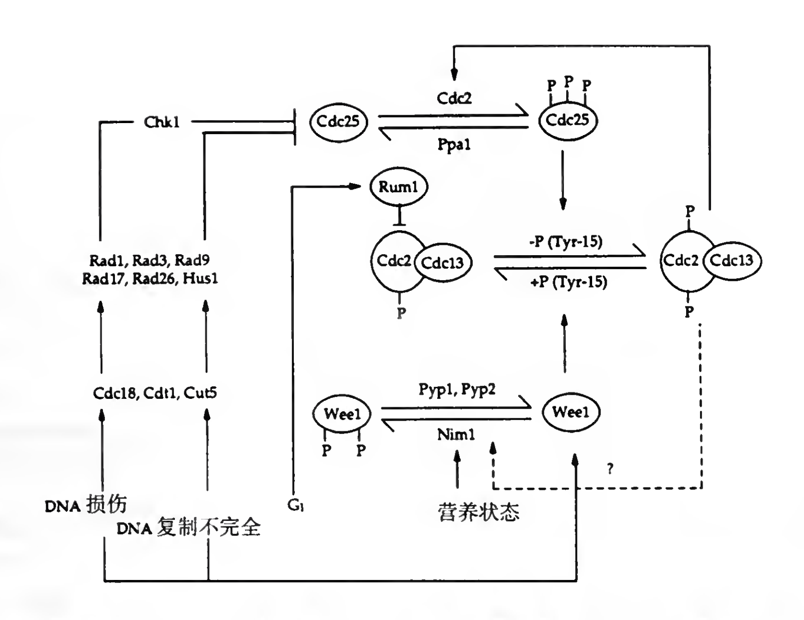


图 2.5 粟酒酵母中进入有丝分裂的调控。进入有丝分裂需要激活 Cdc2/Cdc13 激酶，它们再进一步磷酸化和激活细胞结构改变所需的一系列蛋白。Cdc2/Cdc13 活性主要由 Cdc2 Tyr-15 的磷酸化调控，Cdc25 磷酸酯酶和 Wee1 激酶对 Tyr-15 磷酸化起相反的调节作用。这些调节因子本身由内源和外源信号调节(正如正文中所述，如不完全 DNA 复制、营养状况)。

接来促进分裂期进入中-后期。DNA 损伤关卡可以被 PDS1 蛋白调节，PDS1 控制姐妹染色单体的分离或直接抑制 APC 起作用。有趣的是，巢曲霉(*Aspergillus nidulans*)的 DNA 复制关卡涉及 Cdc2(通过 Wee1 和 Cdc25)和下游成分 NimA 激酶的控制(通过 *bimE* 基因的产物)，BimE 是巢曲霉 APC 的成分。

有丝分裂　M 期的 CDK-cyclin 复合物有丝分裂时使其结构的改变(染色质凝聚、纺锤体组装、细胞核重建和细胞动力学机制的组建)，磷酸化靶蛋白，尽管目前只有个别蛋白被阐明(见表 2.4)。一个重要的但可能不是直接的 M 期激酶的靶蛋白是**后期促进复合物**(anaphase-promoting comples, **APC**)，APC 是破坏盒依赖的泛素连接酶，有助于有丝分裂细胞周期蛋白(在动物中分裂期的 cyclin A 和中期-中后期转换时的 cyclin B——不同的时期可能会反映破坏盒的效率)的快速降解。其他 APC 的靶包括维持姐妹染色单体所需蛋白，可能是着丝粒的成分或是更广泛分布的因子，如上文中所讨论的，酿酒酵母中的一候选者是 PDS1。另一个 M 期激酶的重要靶蛋白是肌球蛋白的轻链，它的磷酸化状态阻止与肌动蛋白的结合和胞质分裂所需细胞骨架环的形成。细胞周期蛋白 B 的降解依赖于有丝分裂纺锤体上染色质的排列，所以 APC 的形成可能依赖于着丝粒的附着并为未附着的着丝粒所抑制。M 期激酶的失活(为 cyclin B 的降解所抑制)允许 M 期激酶的底物如肌球蛋白去磷酸化，这有利于胞质分裂所需的细胞骨架的组织，核纤层重建和染色质去凝聚状态。细胞保证有丝分裂事件按正确的顺序进行，细胞周期作为整体，组织成为一系列相互依赖的事件(表 2.5)。

表 2.4 M 期开始后细胞结构的变化和 M 期激酶的功能

底物	在有丝分裂中的功能
Cdc25	正反馈途径使 M 期激酶活性骤然出现
分裂后期促进复合物(可能是非直接底物)	负反馈途径诱导细胞周期蛋白的降解,引起 M 期起始时激酶活性的瞬时爆发
组蛋白 H1,HMG 蛋白	染色质凝聚
TFIID,TFIIIB,多聚腺嘌呤聚合酶	抑制基本的转录
SBF,SWI5(酿酒酵母)	调节转录(SWI5 的磷酸化抑制细胞核输入)
核纤层蛋白	核纤层蛋白的磷酸化使核纤层蛋白解聚引起破坏核层和驱动核膜溶解
RMSA(有丝分裂纺锤体组建调节因子)	有丝分裂纺锤体组建
波形蛋白 Vimentin	细胞骨架重建
钙调(结合)蛋白(Caldesmon),肌球蛋白轻链	钙调(结合)蛋白是肌动蛋白丝调节因子,肌球蛋白轻链与肌动蛋白相互作用在胞质分裂的位置形成收缩环。它们的磷酸化抑制其活性
Caesein 激酶 IIα 和 β 亚基,c-Src,c-AbI	具有调节作用或其他特异性下游目标的蛋白激酶

表 2.5 有丝分裂不同阶段的重要细胞事件和以楷体表示的分子机制(已知的)

阶段	重要事件
前期 (prophase)	*M 期激酶激活。*染色质凝聚,有丝分裂纺锤体开始形成。这些可能是 M 期激酶直接磷酸化靶蛋白所致。
前中期 (prometaphase)	在后生动物中,核膜破裂允许纺锤体接近染色体。一些纺锤体的微管连接着丝粒。细胞核的重建直接受 M 期激酶的调节。当所有的染色体都附着在纺锤体上,*APC 被 CDC20/Slp1 激活,随之是 MAD/BUB 途径的抑制。*
中期 (metaphase)	Cyclin A 降解。染色体通过附着在着丝粒上的纺锤体,由两极的纺锤体的拉力引导入**中期板** (metaphase plate)。
后期 (anaphase)	*Cyclin B 降解;M 期激酶失活。*配对的着丝粒由于微管的缩短而向两极分离;两极也由于微管的极性分开。负责 cyclin B 降解的蛋白酶解机制可能还触发姐妹染色单体结合的蛋白(Scc1/Mcd1)。这一阶段必须在染色体排列完成后进行。早中期中阻止 cyclin B 降解和过早分裂的信号特点目前还不得而知,但这组成了一个细胞周期中的主要关卡。
末期 (telophase)	*M 期激酶的底物去磷酸化(包括肌球蛋白轻链,核纤层蛋白,组蛋白)。*染色体达到两极,核膜重建,染色质去凝聚,收缩环装配。
胞质分裂 (cytokinesis)	收缩环装配完成。细胞缢缩在残存的纺锤体附近,细胞分裂成两个子细胞。

2.5 动物中特殊的细胞周期系统

从细胞周期中退出 如上文中所讨论的,在 G_1 期的一定阶段中动物细胞在生长因子消失导致停止生长并从细胞周期中退出,进入静止期 G_0。这可能是对生长因子缺乏的瞬时反应,或分化的永久属性(如神经元和肌细胞),或在发育过程中控制生长结构大小。生长因子缺乏使细胞从细胞周期中退出是因为生长因子是 D cyclin 合成所必需的,一些非

常不稳定的蛋白只有在 G_1 早期通过新的转录产生。这些转录由生长因子信号的激活而产生(参见第 28 章信号转导)。简而言之,生长因子是酪氨酸激酶受体的配体,酪氨酸激酶通过 Ras,Raf 和 MAP 激酶启动信号传导通路,最终激活核内的转录因子。这些转录因子的靶蛋白是一系列称为**早早期基因**(immediate early genes)。其中一些编码激活 D cyclin 转录的转录调节因子。G_0 期细胞受生长因子刺激数小时后进入 S 期,反映了 D cyclin mRNA 在细胞质中出现的时间。

从细胞周期中退出同样需要生长抑制因子如 TGF-β 的促进,接触抑制信号或失去底物。这些分散的信号通过小分子细胞周期依赖激酶的抑制剂(CKIs)作用特异触发细胞周期机器的成分。抑制剂家族 p16 通过防止 CDK-D cyclin 组装和对 RB-1(参见*肿瘤抑制基因*)蛋白的磷酸化能力特异性抑制其活性。p21 和 p27 家族抑制所有的 CDK-cyclin 活性使细胞周期阻滞,它们通过抑制 CAK(本身是 CDK-cyclin 全酶)对 CDK-cyclin D 复合物的激活达到目的,同时还抑制了含 cyclin E 和 cyclin A 复合物和 PCNA/δ-DNA 聚合酶复合物本身的活性。CKI 也汇集内源性途径的信号。DNA 损伤的主要调节途径通过激活 p21 的 p53 调节因子起作用(参见第 18 章癌基因和癌)。

与刺激 D cyclin 合成一样生长因子抑制许多上述 CKI 以促进细胞生长。具有讽刺意味的是,发现 p21 或 p27 在细胞中的始终存在,并是正常有功能的 CDK-cyclin 复合物的一部分。抑制蛋白的化学剂量的结合并不影响激酶活性,而是在高浓度时抑制作用才表现出来。p21 也能直接与 PCDA/DNA 聚合酶 δ 亚基结合,DNA 损伤抑制 DNA 复制比通过 CDK-依赖激酶的信号传导通路更直接。

细胞凋亡是细胞周期阻滞的另一种方式　**细胞凋亡**(apoptosis)又称为**细胞程序化死亡**(programed cell death),是细胞对不同的内源和外源信号反应后启动细胞死亡的一种形式[和坏死(necrosis)相反,后者是由于损伤、感染和伤害所引起的]。细胞凋亡在发育中起重要作用,也是对化学伤害和感染的抵御。抑制细胞凋亡可以引起癌症。

凋亡中的细胞的行为与坏死截然不同。它涉及染色质的凝聚和外周化、细胞质的减少,核的片段化、细胞器的致密化,与周围细胞联系的中断、内质网与细胞膜的融合和最终细胞片段化形成很多细胞**凋亡体**(apoptotic body),被周围细胞吞入。没有像细胞坏死中所见到的发炎现象。染色质凝聚之前,DNA 降解成存在多个核小体单位的小片段。

控制细胞凋亡的重要细胞死亡调节因子是 BCL-2 家族。BCL-2 本身是**存活因子**(survival factor, 即它抑制细胞凋亡),与在线虫(*C. elegans*)发育过程中存活细胞表达的 CED-9 蛋白同源。家族中的其他成员(如 BAX,BAD 和 BAK)是细胞凋亡的刺激因子。这些蛋白形成二聚体,与其竞争,此反映了它们的相对丰度,这种竞争决定了细胞的命运。BCL-2 信号的效应因子是与白介素-1β 转化酶(**ICE 蛋白酶**)相关的半胱氨酸蛋白酶家族。很多这样的蛋白酶在哺乳动物细胞中以无活性的酶原的形式合成,通过一系列蛋白酶解激活,从而使如多聚 ADP 核糖聚合酶和核纤层等细胞凋亡的直接效应分子的活性达到最高。

引起细胞凋亡的信号传导途径还未被完全阐明。其中涉及 p53(参见第 18 章癌基因和癌)的是在某些细胞类型中(如皮肤细胞)对 DNA 损伤启动细胞凋亡是抑制 BCL-2 和激活 BAX。在其他一些细胞中,p53 通过激活抑制 CDK-cyclin 复合物和应激-激活蛋白

激酶(stress-activated protein kinases)(参阅)的 CKI,使细胞阻滞在 G_1 期。这些途径在发育过程(如肢芽的趾间区)中并不起作用。因为小鼠中 p53 的同源基因 *TP53* 缺失后发育完全正常。

其他信号,如生长抑制因子 TGF-β 也可诱导细胞凋亡,它是通过 BCL-2 家族途径但不依赖于 p53。也有直接影响 ICE 家族蛋白酶的信号途径。如肿瘤坏死因子(TNF)通过胞浆中 80 个残基**死亡结构域**(death domain)启动细胞凋亡,这结构域为 TNF 受体和其他一些受体、以及胞内凋亡信号分子所共有(如 TRADD,Fas,MORT1 和 RIP),这些分子都可以激活 ICE 蛋白酶。CrmA 是牛痘病毒编码的 ICE 蛋白酶的抑制剂。这个存活因子可以阻断由 p53、TGF-β 和 TNF 所引起的凋亡,使病毒在感染后的免于死亡。

(徐　沁 译)

第3章
染　色　质

基本概念和定义

- 真核细胞染色体中的 DNA 通常是和质量较为均衡的蛋白质一起以复合物的形式存在,形成高度有序的核蛋白物质,称为**染色质**(chromatin)。
- 染色质中的蛋白质主要成分主要由**组蛋白**(histone)构成;此外尚含有少量其他蛋白,统称为**非组蛋白**(nonhistone)。组蛋白相对比较均一,以维持染色质结构的基本形状;而非组蛋白是不均一的且执行多种不同的功能。每条染色体中的 DNA 则是单个很长的分子[称之为**单线模型**(unineme model)]。
- 一个真核细胞核的直径不超过 1μm,但是它含有 10^9bp 以上的 DNA,这样大小的 DNA 分子如果完全延展开来长度将达到几米。因此,这种 DNA 必需是一种高度折叠的,这种性质表述为 DNA 的**包装比**(packaging ratio),即细胞核内 DNA 的长度与其延展状态的理论长度的比值。间期核中 DNA 的包装比约为 10^3。
- DNA 包装成染色质成分的过程是高度有序化的,经过了一系列不同级别的结构次序来完成。染色质不仅仅是一种物理结构,它还是影响基因活性的重要因素之一。染色质的组成调节基因表达,并是后生性细胞记忆的基础。

3.1　核小体

核小体结构　　**核小体**(nucleosome)是染色质的基本结构单位,几乎在所有真核生物中都有存在,除了双鞭甲藻(dinoflagellate)以外,结构保守。当细胞核在低盐溶液中裂解时,染色质解聚、核小体以串珠状的形式释放出来。

核小体与 DNA 之间的结构关系可以通过核酸酶的作用加以阐明(图 3.1)。解聚后的染色质经过核酸酶的部分消化分离,成为由**单个核小体单位**(mononucleosome)或多联核小体组成的短小片段。与单个核小体结合的 DNA 长度是恒定的,一般在 200bp 左右(精确长度具有种属和细胞类型的特异性)。经过消化裂解后的染色质所释放出的 DNA 通过电泳分离,可以呈现出与单个核小体 DNA 长度的整数倍的阶梯状特征谱图,表明核酸酶的剪切作用位点是在每个核小体之间相对恒定的位置上。

在大多数物种中,核小体 DNA 在核酸酶的进一步消化作用下其长度可以缩短为 165bp。此过程是通过一步反应完成的,提示在最初的核酸酶剪切作用之后,紧随其后的是 DNA 游离的末端被快速修剪。保留在核小体上的 DNA -受到核小体颗粒蛋白的保

护，被称为**染色小体**(chromatosome)。

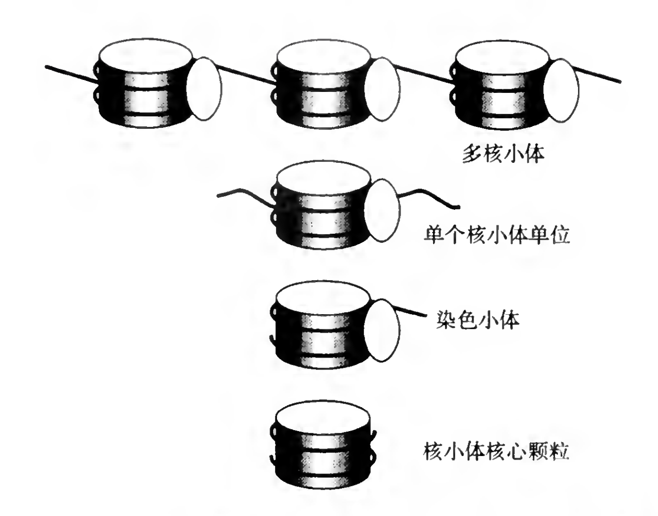


图 3.1 染色质中核小体的结构。释放的 DNA 以串珠形式存在，串珠由核小体通过连接 DNA 连接而成，用微球菌核酸酶来剪切，产生单个核小体。经进一步消解切除大部分连接 DNA，产生单个的染色小体。经进一步消化，与 H1 组蛋白结合的 DNA 也被切除，解离出核小体核心颗粒，它含有余下的 146bp 的 DNA。

在核酸酶的继续消化作用下，核小体中的 DNA 长度减小至 146bp 左右，同时有少量的蛋白质成分也从核小体中释放了出来。这样长的核小体 DNA 在各种真核细胞中都有存在，并且具有一定耐受核酸酶作用的能力。核小体 DNA 的存在形式与组成核小体核心部分的蛋白质关系密切。**核小体核心颗粒**(nucleosome core particle)由组蛋白八聚体和 DNA 组成，四种组蛋白 H2A、H2B、H3 和 H4 每种 2 个拷贝组成八聚体结构，长度为 146bp 的核心 DNA 在组蛋白八聚体外环绕将近 2 周。核小体核心颗粒之间由**连接 DNA**(linker DNA)相连，连接 DNA 的长度在不同种属和细胞类型之间差异较大。对于大多数的物种而言，DNA 进出核心颗粒上 DNA 的进端和出端与**连接组蛋白**(linker histone)——组蛋白 H1 相结合，在此处 H1 将 DNA 的两端封闭在一起。

核酸酶对染色质的消化作用可概括如下：核酸酶最初在连接 DNA 的特殊位点剪切，然后修剪 DNA 末端，直到受到组蛋白保护，而保留在核心颗粒上。在核酸酶的进一步消解下，与 H1 组蛋白结合的 DNA 也被切除，H1 的丢失表明 H1 组蛋白位于核心颗粒以外，而紧靠着它。经过核酸酶作用后残余的 DNA 被核小体核心颗粒保护起来。在酵母染色质中似乎没有连接组蛋白，因此在核酸酶的最初消化作用下，产生的单个核小体 DNA 的长度为 164bp，然后经过一步修剪成为 146bp 的核心 DNA。连接核小体核心颗粒之间的 DNA 长度大约为 18bp，是已知最短的连接 DNA，然而在酵母的基因组中组蛋白 H1 基因已被鉴定出来。

组蛋白　组蛋白是高度碱性的蛋白质(富含带正电荷的赖氨酸和精氨酸残基)，它折叠形成一个带有凸出 N 端尾巴的致密核心。这些正电荷的残基，可以与 DNA 相互作用，DNA 通过携带负电荷的 DNA 骨架与蛋白形成盐桥连接。而组蛋白游离的尾部是翻译后

修饰的靶、并能促进核小体间相互作用以及与其他染色质蛋白质之间的相互作用。目前已经有几百个组蛋白编码顺序被测定，显示出在真核生物中具有明显的保守性。组蛋白H3和H4保守性最强，各种连接组蛋白则显示出很大的多样性。除了H4组蛋白以外，所有的组蛋白都是以多种同等型的**异型组蛋白**(isohistone)存在的，这种现象与染色质的**超优势度**(predominance)在不同细胞类型中以特异方式改变有关。连接组蛋白显示出很多的异型多样性，而且分成不同亚类(H1、H5、H10)。

组蛋白具有特定结合形成各种复合体的固有能力，而且不需要DNA的参与，比较典型的是$(H3-H4)_2$四聚体和H2A-H2B二聚体。每个组蛋白都含有一个三α螺旋的基序，称为**组蛋白折叠**(histone fold)，它可促进相互作用。体外实验证实，组蛋白四聚体可以与DNA组装成核小体样的颗粒，同时加入其他组蛋白，染色质就可以自动组装，该过程通常是缓慢进行的；而在体内，核小体的装配还需要有**分子伴侣**(molecular chaperone)的参与(参阅)。用X射线晶体衍射以及化学交联(cross-linking)对组蛋白的八聚体进行了研究，目前分辨率已经达到2.8Å。H2A-H2B二聚体存在于中心$(H3-H4)_2$四聚体的上下部位，由此形成圆柱状颗粒的外表面DNA在其上超螺旋附着的地方，此表面是光滑的，由于很多弯曲存在使表面变成曲折不平。当DNA进出组蛋白四聚体时，会与H3组蛋白的游离末端结合。

所有的组蛋白似乎都要经历翻译后修饰的过程，这种修饰作用通常发生在NH_2-末端，并且在大多数情况下，修饰的特殊方式与染色质功能的变化相关。H3和H4组蛋白的**乙酰化**是易接近的“开放”染色质标志之一，但最近也有很多实验表明乙酰化与遗传活性之间的关系并不是简单的。H3和H1组蛋白被磷酸化[至少在黏菌(*phsyarum*)是这样的]，H1组蛋白的磷酸化程度是按细胞周期而变化。各种组蛋白也会被甲基化、或者与泛素(ubiquitin)或多聚ADP核糖(poly-ADP-ribose)结合，但这种修饰作用的意义目前还不十分清楚。组蛋白的尾巴长而未形成一定结构，这对于核小体之间的相互作用是必需的。因此组蛋白的修饰被认为是瓦解核小体之间的相互作用，并扰乱了高度有序的染色质。

非组蛋白　在染色质中的蛋白成分里，组蛋白占有相当大的比例，而且，在性质上相对较为均一。染色质中其他的蛋白质统称为**非组蛋白**(nonhistone protein)，虽然含量很少，但表现出极端的不均一性。非组蛋白主要包括与DNA和组蛋白的代谢、复制、重组、转录调控等(参阅)密切相关的各种酶类。它们还包括形成染色质高级结构的**支架蛋白**(scaffold protein)和**高迁移率蛋白**(high mobility group protein，HMG protein)，这类蛋白携带较高的电荷，在染色质结构组成以及基因调控中发挥各种各样的作用。其中核小体结合蛋白的HMG14/17家族可积聚于活化的染色质上，推测这一现象有助于染色质较高级结构的解聚。HMG I/Y家族的非组蛋白它们的确切作用还不十分清楚，但已知其优先与含有AT重复序列的DNA相结合；并且，和H1型组蛋白类似，它们可以被周期蛋白依赖性激酶的激酶所磷酸化(参见细胞周期章节的有关内容)。在包装和转录激活等过程中，HMG蛋白的另一个重要作用是使DNA产生角度较小的弯曲(sharp bend)，在包装的过程中DNA需要采用这种特殊的三维构象，而在转录过程中，在其相邻的不同位点上将结合调节因子(参见SRY因子，增强体)。还有一类非组蛋白是**鱼精蛋白**(protamine)，它可以促进DNA包装进精子头部。这类蛋白排列在DNA双链相邻的大沟中，使DNA折叠

成高度致密的一束平行纤维结构。

组蛋白-DNA 对接和连接数的矛盾 在 DNase I 或者自由基的作用下，核小体核心颗粒中的 DNA 可降解为 10～11bp 大小的片段，这一现象提示：缠绕于组蛋白八聚体外的 DNA 是以 B 型构象的形式存在的(参见核酸结构等章节的有关内容)。上述这种剪切作用可以发生在核小体表面的各种部位，但多发生在其两端的位置，而很少发生于核小体的中部，这说明当 DNA 缠绕于八聚体组蛋白的外表面时，其结构发生了某种变化。DNA 在核心颗粒外缠绕 2 圈，然后形成近 1.8 周的负超螺旋结构。通过 DNA 结合成核小体从而形成超螺旋程度的检测实验结果则表明，每个核小体实际只产生正好 1 周的负超螺旋。此矛盾称为**连接数矛盾现象**(linking number paradox，参见 DNA 拓扑学)。对上述现象的一种可能的解释，是由于核小体中 DNA 的螺距(每周螺旋含有的碱基数目)，比游离 DNA 的要小一些的缘故。

核小体与 DNA 结合的位置 核小体与 DNA 不是随机结合的，而是有优先结合位点。这种特性称为**核小体分相**(nucleosome phasing)，并可以通过微球菌核酸酶和限制性内切酶对染色质的消化实验来证实。上述实验将产生不连续的 DNA 片段，这是因为相对于每条染色质中的核小体而言，DNA 内部的限制性酶切位点总是相对一致的。核小体分相通常以两种方式出现：或者是通过特殊顺序上的组蛋白八聚体使 DNA 弯曲而定位；或者是通过核小体在特殊界面受到排斥而定位，这样可获得两类分相。产生弯曲的特殊 DNA 序列有影响核小体的定位的趋势(譬如启动和增强子部分)，正如其他一些蛋白质对无核小体区域，如启动子和增强子区的结合。对于酵母染色质而言，核小体分相发生于转录阻抑的染色质上；而在活化的染色质上，核小体的结合位置是随机的。这一现象提示：核小体分相可能是起始染色质高级结构的形成所需要。

3.2 染色质的高级结构

30nm 染色质丝 DNA 盘绕成核小体仅仅是形成染色质的一级结构。当细胞核放在低盐的溶液时，染色质是以典型的念珠状形式存在的，即通常所谓的 **10nm 染色质丝**(10nm fiber)，此时的 DNA 包装比约为 5，且没有 H1 型组蛋白。而在较高浓度的盐溶液中，染色质以更致密的形式存在，并需要在 H1 组蛋白，形成直径 25～45nm 的螺线管样结构，即为 **30nm 染色质丝**(30nm fiber)。虽然也有人提出另一种模型：如“Z”字型结构；或者核小体并不形成规则结构，而是以不规则的团块或者**超颗粒**(superbead)以线性方式排列，形成纤维。

人们根据 30nm 染色质丝的密度(30nm 染色质丝是由 10nm 的念珠状染色质纤维盘绕形成的，每圈含有 6～8 个核小体，而 DNA 的包装比增大为 40)，对其结构提出了多种不同的结构假说，但根据 X 射线衍射实验结果表明 30nm 染色质丝中的核小体盘的平面平行与螺旋轴成一定倾斜角。目前关于连接区 DNA(linker DNA)的功能和组蛋白 H1 的位置争论较多，它们可能是通过协同结合的方式形成极性丝状体的。

染色质环 处于有丝分裂间期的细胞核中染色质的整体结构并不十分清楚，但推测

30nm 染色质丝附着在**核基质**[1](nuclear matrix)上的不同位点,形成一系列的环状结构,内含 30~100kb 大小不等的 DNA[即**折叠纤维模型**(folded fiber model)]。去除了蛋白质成分的染色质通过扫描电子显微镜可以清楚地观察到这种环状结构,它们碱基与核基质中的支架蛋白结合(参见灯刷染色体)。这种环状结构的功能学意义不清,但它们与以后将要讨论的染色质功能区(chromatin domain)有关,它们已经被遗传学和生物化学的分析所鉴别。

常染色质和异染色质　　有丝分裂间期的染色质以两种截然不同的形式存在:一种是较为松散的**常染色质**(euchromatin),正如前面所述的,它主要由 30nm 染色质丝缠绕、混合而成,多存在于核质中;另一种是高度浓缩的**异染色质**(heterochromatin),以高度有序的结构形式存在于细胞核的周边部位。

异染色质通常处于转录抑制状态,据推测它们高度有序的结构与有丝分裂期的染色质相似,不含有转录调控蛋白。尽管 DNA 序列不是很重要的决定因素,目前对于异染色质的结构特性仍知之甚少,譬如**组成型异染色质**(constitutive heterochromatin)(在各种细胞中任何时期都是以浓缩的形式存在,并且经常位于着丝粒附近)。真核细胞中还含有**兼性异染色质**(facultative heteromatin),它在某些细胞中是保持着抑制状态;而在另外一些细胞中则恰好相反,例如失活的 X 染色体。似乎一些特殊的核蛋白和调节异染色质的结构有关,从几种不同的组织中已分离出可能具有上述功能的候选者,它们都具有调控与花斑位置效应有关的现象(见下文)。高等真核细胞的异染色质还含有比例较高的经过特殊修饰的组蛋白,这种现象与 DNA 甲基化水平增加有关。

中期染色体　　有丝分裂中期的染色质处于高度凝聚的状态,此时的染色体被包装为微小的、易于分离的独立结构。关于染色质高凝聚状态(包装率 10^5)的形成同样也不是十分清楚。但已知这种高度浓缩的染色质丝以环状结构附着于蛋白骨架之上,形成由中心螺旋发出的染色质放射状环。尽管还不知道这些环状结构与前面讨论过的分裂间期细胞核中的染色质环是否属于同样的结构,但这两种结构形成所需的 DNA 基体是相似的。骨架结构存在于每一对姊妹染色单体上的相应部位,并且在有丝分裂开始时彼此连接起来。

有丝分裂期的染色质最显著的特点是它虽然处于高度浓缩的状态,但仍然可以保持在以前分裂间期时何种基因具有转录活性、何种被抑制的记忆能力。同时,在有丝分裂期染色体中 DNA 的包装也不尽相同,这样的组织结构为随后的分裂间期染色质去凝聚状态时,其开放或关闭染色质结构域的特殊排列得到恢复。有丝分裂染色质 DNA 的不同包装,可以通过不连续的染色体显带(chromosome banding)技术(参阅)来显示。利用该技术处理后的染色体在显微镜下显示不同形式的明暗区带,相应区域的遗传学活性可以通过生物化学分析来决定。

---

1) **核基质**[nuclear matrix,也称为**核骨架**(nuclear scaffold)],是细胞核经过去垢剂、核酸酶和高浓度盐溶液抽提后所能残留的结构成分。它组成纤维网格状结构,将细胞核内隔成许多独立的区域,且组成单个的染色体。核基质的成分还不十分清楚,已知其中含有拓扑异构酶Ⅱ(topoisomerase Ⅱ),这种酶也是中期染色体的**中心骨架**(central scaffold)、或者染色体**核心颗粒**(chromosome core)的一种成分。目前虽然已经有一些证据表明了核基质与 DNA 功能关系,但是尚缺乏充分的证据证实染色质与核基质相互作用在 DNA 复制和转录起始等过程中起作用。

3.3 染色质和染色体的功能

染色质和从 DNA 中获取的信息 DNA 组成了高度有序的结构(在这种结构当中,其全长都伴随有蛋白质的存在),因此而引出的问题就是:究竟何种蛋白质介导 DNA 的功能。尤为特别的是复制和转录必须在核小体区域进行,同时需要大量的酶复合物参与,这些酶复合物沿 DNA 方向持续移位,但又不围着 DNA 盘绕。DNA 和 RNA 聚合酶都比核小体大许多。当 DNA 组成染色质时,是如何从 DNA 中获取信息将是重要的。

在 DNA 复制时核小体的结构 在 DNA 复制的过程中,在复制叉中核酸酶敏感性特殊模式在复制中亲代双链的分离表明 DNA 的组蛋白被取代了,游离的组蛋白很快会与子代 DNA 双链重新缔合。在体外实验中当组蛋白通过复制很快在裸露 DNA 上竞争性装配成核小体。用扫描电镜观察复制的 SV40 DNA 证实了这个过程,显示核小体是以念珠状存在于亲代链和两条子代双链上很快向复制叉靠近。

染色质的复制是分散的,而不是保守的。从亲代双螺旋上解离下的组蛋白表明并不优先结合任何一条子代的 DNA 双链,而且还会与细胞周期中 G_1 期新合成积累下来的组蛋白混合(参见细胞周期)。在体内的精细装配机制还不十分清楚,但一种称为 N1/N2 的分子伴侣可以通过将$(H3\text{-}H4)_2$四聚体装到 DNA 上而启动核小体的形成,与之相对应的另一种分子伴侣核质蛋白(nucleoplasmin)则可以促进 H2A－H2B 组蛋白二聚体的停靠作用。关于染色质功能人们提出另一个问题是通过成功的复制,染色质中的活化和抑制区域是怎样稳定地传递下来,这似乎反映了伴随着复制叉的通过,经特殊修饰的亲代组蛋白立即分配到两条子链上,维持了原有染色质的结构。在复制过程中,无论是核小体还是转录复合物,都可以竞争占据 DNA 上重要的位点,这样,就为细胞的变化趋向提供不同的选择机遇,特别是在复制过程起始之前,如果有新的转录调控因子合成的情形下,这种机制就显得比较有意义了,竞争作用的结果将会反映细胞内调控复合物与 DNA 之间的亲合能力及其在细胞中的有效浓度(参见下文)。

转录过程中的核小体结构 和复制过程一样,转录开始时核小体与 DNA 分离,而紧随 RNA 聚合酶后核小体又重新装配起来。大多数可转录的基因都保持有核小体的结构,而非转录基因的核小体分相特殊模式不复存在,结果经微球菌核酸酶和限制性内切酶消化后,DNA 电泳显现为界限不清晰的成片条带,而不是不连续的条带。检测聚合酶复合物在核小体中作用的实验表明:当酶复合物通过核心颗粒 DNA 一半位置的时候会暂停,这似乎说明当酶试图通过核小体释放的第一个螺旋时,染色质中存在着一种扭应力。当酶通过这一位点后张力被释放,八聚体组蛋白被解离出来。在酶作用之后,八聚体又与 DNA 重新结合,多半是由于八聚体可能是与非转录链结合;也可能是它与酶本身暂时结合。

转录水平很高的基因,如**灯刷染色体**(lampbrush chromosome)(参阅)的 rRNA 基因,在其染色质的构象是伸展的,表明没有核小体的结构存在。转录延伸复合物后面的转录后的 DNA 不能与解离的组蛋白重新装配。而在非常活跃的基因中构成的染色质中,存在着不定数的无核小体区域,从而能保护 RNA 聚合酶正常发挥作用。

染色质结构域 具有转录活性或潜在活性的染色质其结构特性与无活性染色质是截然

不同的。具备转录能力的染色质，对 DNase I 的消化作用普遍较为敏感，说明它们是以相对较为松散的结构包装的，因此又称为**开放染色质**(open chromatin)。事实上，转录作用对于核小体的结构也具有破坏作用，而染色质 DNA 对 DNase I 的敏感区域可以延伸到实际转录单位两侧几千个碱基外，并且这种敏感性在转录没有开始之前就已经存在了，提示这种现象涉及到比核小体更高级的组织结构。

对 DNase I 敏感性的程度确定了**染色质结构域**(chromatin domain)的概念：即在这样一个特定的染色质区域当中，其活性与染色质其他部分无关。对 **DNase I 普遍敏感**(general DNase I sensitivity)的区域统称为 **DNase I 高敏感区**(DNase I hypersensitive site)。这些结构区通存在于转录单位的两侧，并且能优先被较低浓度的 DNase I 所剪切消化。它们的碱基长度大多在 200bp 左右，并且与顺式作用[参见调控元件、增强子(enchancer)、基因座控制区(locus control region)]有关。它们的存在将有利于染色质提供没有核小体结合的区域，使得转录调控复合物可以与 DNA 有效地结合。上述论点可以通过对 SV40 基因组的研究所证实，DNase I 水解位点的分子生物学分析表明与调控元件相关；通过扫描电镜进行的拓扑学研究亦已鉴别了无核小体结合区。虽然开放染色质结构的分子基础仍未完全被阐明，但已经有了一些有趣的相关性，开放染色质含有比其他大部分染色质更多的氨基端乙酰化的核心颗粒组蛋白和 HMG 14/21 蛋白，并且连接组蛋白的含量相对较低。由于连接区组蛋白对于 30nm 染色质丝是必须的，并且组蛋白 N 端尾巴可促进核小体之间相互作用，所以处于抑制状态的染色质转化为开放染色质的有可能和 30nm 染色质丝解聚成为简单 10nm 染色质丝结构有关。这时的染色质主要依靠非组蛋白来稳定。组蛋白乙酰转移酶和脱乙酰酶显示可能会被一些转录因子所招募，提供了**染色质重组机制**(chromatin remodeling)，成为转录活化或抑制过程的起始步骤。在哺乳动物和植物之中，**抑制型染色质**(repressed chromatin)通常与 DNA 高度甲基化有关。而 DNA 的甲基化在基因表达的调控方面可能起主要作用(参见 DNA 甲基化)，而 DNA 甲基化在许多低等真核生物细胞中是不存在的，如酵母和果蝇中。

真核生物基因组可以被分成功能各异的独立区域，这种结构方式对于哺乳动物而言，具有显著的意义，因为染色质的结构与 M 期的染色体在拓扑学上是相互对应的(参见染色体的结构与功能)。采用各种不同的分化染色方法(参见常见的如 G-带和 R-带显示技术)，可以重复地显示明暗相间的横向带谱，与染色质结构的不同的密度区域相对应。G-带与其他方法显示的染色体带是一致的，例如：瞬时复制显带(transient replication，此技术可以特异性地标记早期和晚期复制的 DNA)；和 D-带技术，此技术可以鉴别 DNase I 敏感区域。分子生物学的实验证据表明：这些区域中的 DNA 序列与其结构组成之间存在着差异，这就为哺乳动物基因组的双相组织结构(biphasic organization)(参阅)提供了一定的理论依据。

功能区及其边缘部位的功能　　染色质在物理学上可以划分成许多不连续的结构，从拓扑学上也可分为独立的区域。此已被染色质上去除了蛋白质的环和骨架结构所证实；灯刷染色体(lampbrush chromosome)(参阅)的侧环；以及果蝇多线染色体(*Drosophila* polytene chromosome)可重现的区带图谱等所证实。染色质环的基部系于核基质上，似乎是和特殊的核蛋白复合物有关。染色质环的意义还不十分清楚，推测它们可能与前面讨论过的独立功能区域的形成有关，正因为如此，染色质上相邻区域可以采取不同级的包装形

式，并且可能为**增强子单配性**(enhancer monogamy)提供一种机制，即每一个特定的增强子都有其限定的靶基因。

如果染色质环能够与其功能区很好地对应，就可以预测哪些 DNA 序列与核基质具有相互作用，这些序列应当分布于染色质功能区域的边缘部位，从而就能从效应区域的邻近部位将有关基因分离出来。它们很可能是以**绝缘子**(insulator)或者**边界元件**(boundary element)等形式存在的。

曾经推测存在的**基质结合区域**(matrix-associated region, MAR)[亦称为**核骨架附着区**(scaffold attachment region, SAR)]，已经根据其结合核基质蛋白的能力，通过两种互补的实验方法被鉴别出来。在不溶性的成分中 DNA 片段与基质成分接触能捕获到假定的 MARs。另外，去除了蛋白质成分的染色质经过核酸酶消化以后，仅有受到 MARs 保护的部分，不受核酸酶活性的作用。上述这些实验还鉴别了一些富含 AT 碱基的元件，除了具有拓扑异构酶 II 的识别位点以外，它们没有很严格的序列保守性。此外还鉴别出核基质以及分裂中期的核骨架两种成分。很有可能拓扑异构酶 II 将染色质分成为拓扑学隔离功能区。

通过细胞生物学和生物化学等方法也鉴别了一些边界元件，它们一般定位于已知染色质结构区的边缘部位，并含有核酸酶高度敏感的位点。这类元件包括果蝇特殊染色质结构(*Drosophila* special chromatin structures, scs 和 scs′)以及鸡和人的来自**基因座控制区**(locus control region)(参阅)的 β-珠蛋白 HS5 元件。而一个真正完整的边界元件应该能以自主方式建立一个独立染色质功能区域，而 β-珠蛋白的座位控制区域(locus control region, LCR)在体内的作用被认为具有这种特点。这种功能可以通过两种方法检测，一是经检测旁侧边界元件可保护转基因不受内源性位置效应的作用；二是边界元件经检测可使增强子产生“绝缘”。用上述方法已经成功地鉴定了果蝇的 scs 样元件、鸡 β-珠蛋白 LCR 的 5HS 位点以及鸡溶菌酶基因 A 元件。这些元件之间的可交换性存在着一种梯度，例如：β-珠蛋白 LCR 5HS 位点在果蝇是有功能的，而 scs 样元件在转基因小鼠中就没有功能；鸡的溶菌酶 A 元件在转基因鼠和转基因植物中都抑制位置效应[也可参阅吉普赛转座子(gypsy transposon)]。

尽管对用物理方法分离的 MARs 功能进行检测实验的结构并不明确，但某些边界元件的功能鉴别表明染色质物理结构分成为很多环，此可能是划分染色质功能区域的基础。

异染色质结构和外遗传基因调控 常染色质 DNA 易位至异染色质区通常会导致转录抑制播散到常染色质区域内。在不同的细胞中异染色质化范围是不同的，但是以克隆方式扩增的，因此，染色体断裂点附近的基因表达形式就多种多样了。**花斑位置效应**(position effect variegation, PEV)这种现象表明，顺式作用潜在的活性可以穿越染色体断裂点。这种播散效应是线性的且不会间断，因此染色体断裂点附近的基因座虽然处于抑制状态，但这种抑制状态从不会跳过基因。这就提示异染色质的正常传播过程是通过特殊的蛋白加到业已存在的异染色质上，然后逐渐扩大范围，直到遇到某个边界元件为止。易位的过程可能移去了这些边界元件，可变化的播散导致产生了 PEV 现象，此反映出在不同的细胞当中，与异染色质相关的蛋白质总数是可变的。

异染色质蛋白和常染色质蛋白似乎是均衡地存在于核质和染色质之中的，带有染色质功能的一些突变编码基因将改变这些蛋白的平衡状态，而伴随着易位事件导致 PEV 的

范围发生位移[这种情形称为**完全相反效应**(antipodal effect)]。在通常情况下,上述这种突变其行为或将成为 PEV 的抑制子或成为 PEV 的增强子,和染色质高级结构的基因将可被鉴别。

在果蝇中 PEV 调节物中寻求那些各种细胞中已分离出在异染色质中起作用的成分。异染色质本身是最好的 PEV 调节物之一,通过增加 Y 染色体的量(大多数为异染色质),PEV 也可以被抑制,这可能是由于核质中异染色质蛋白减少的缘故,其中组蛋白也被作为一种重要的成分被鉴别出来。异染色质通常较少含有乙酰化的 H4 型组蛋白,但富含其特殊乙酰化的形式。还有一些调节物与在异染色质结构中发挥直接的作用的特殊蛋白相关,包括通用蛋白质 HP1,以及某些基因表达的调节物,如多栉状体(polycomb,参见同源异型基因、分化的维持),它们在染色质结构水平上介导它们的作用。这些蛋白质常有一个保守的结构,称为**染色质功能区**(chromodomain)。

在酵母细胞中,异染色质蛋白根据沉默交配型基因座去阻遏突变而被分离出来。此基因座常隐藏在端粒附近的抑制性染色质[参见交配型转换(mating type switching)],和组蛋白 H3、H4 类似的沉默信息调控因子 SIR3、SIR4 以及另外一种称为抑制-激活 RAP1 的蛋白质都已经被分离。很明显,异染色质化的扩展反映组蛋白和 SIR 产物之间的相互作用、并被 RAP 所促进(如 SIR3 与组蛋白 H4 的 N-末端尾巴相互作用)。

3.4 细菌类核的分子结构

细菌染色体的组成 细菌细胞含有单个的染色体,它和真核细胞染色体一样,是以核蛋白复合物的形式存在的。这种**类核**(nucleoid)在细胞分裂之前是以不凝聚的状态存在的,也不呈现真核染色体结构致密的特点,但它也会形成一系列的环状区域,就像反应了结构有序性一样也反应了其不同的功能。

在大肠杆菌(*E. coli*)中,上述这种环状结构区域的长度大约在 50kb 左右,在整个基因组内有接近 100 个环。每个结构区域在拓扑学上是相互独立的,这样就有利于对不同的启动子进行独立的拓扑控制。这种组成的分子机制目前还不清楚,例如:尚不知道这些环状结构是限定于染色体的特定区域,还是随机形成;还未弄清这些结构与染色体内何种序列有关;而且也不知道这些环状结构是如何形成与维持。类核中一系列成分已被确定(如:HU、H1),虽然影响到其他系统发生变化的一些基因的突变已被鉴定出来,但影响类核功能的突变尚未被鉴别,表明类核蛋白可能是较为丰余的,并且可以间接地影响到其他的系统,例如:可能会通过修饰 DNA 的拓扑学状态,从而作为一种重要的转录激活调控因子(参见第 29 章转录,第 16 章核酸结构)。

(方向东　王　琪译)

第4章
染色体突变

基本概念与定义

- **染色体突变**[chromosome mutation,或称**染色体畸变**(chromosome aberration)]是涉及基因组大片断的突变。这样的突变往往影响很多基因并且可在细胞遗传学水平上观察到,即用光学显微镜可见。
- 染色体突变如包含染色体数目变化的**数量**(numerical)突变,和涉及染色体片断的断裂与重排就是**结构**(structural)突变。数量突变往往是由结构突变导致的染色体分离错误引起的,但也可能反映了异常复制或受精时的错误。结构突变是由断裂染色体的错误修复或非等位重组事件引起的。在哺乳动物与果蝇中,微细的结构突变也可被监测到,因为它们破坏了*染色体带型*(参阅)。如果 DNA 只发生重排,而没有物质的丢失与获得,这种结构突变是**平衡**的(balanced);如发生了 DNA 的丢失与获得则是**不平衡**的(unbalanced)。所有的数量突变都是不平衡的。
- 染色体突变有四种后果,断裂、融合、位置与剂量效应。**断裂与融合效应**(disruption and fusion effect)发生在结构突变中并且反映了突变前与突变后染色体断裂位点的性质(例如一断裂位点将一基因与其启动子断裂或分离,导致基因功能的丧失,而融合将两个基因连接到一起,导致一复合产物的合成,可能具有新的功能)。**位置效应**(position effect)也发生在染色体结构突变中,反映了染色质结构对基因表达的总体影响。(例如,转座可能会使一正常活性基因与一异染色质区域相邻,从而抑制转录,参见*染色质区域,花斑位置效应*)。剂量效应在结构与数量突变中都有发生,与细胞内基因的拷贝数有关。很多常染色体基因的剂量水平是比较灵活的,但另一些基因表现出*单倍不足性*(haploinsufficiency)(参阅)这类剂量效应。不平衡的染色体突变,特别是完全的非整倍性,同时改变了很多基因的剂量,因此往往增大了剂量效应。很多连续的特别是整条染色体上的所有基因的剂量的改变,被称为**染色体不平衡**(chromosome imbalance)。
- **组成型染色体突变**(consititutional chromosome mutation)发生在生殖系或合子中,因而影响身体中的每一个细胞。在人类中,这会引起特异的临床综合征[例如,唐氏综合征(Down's syndrome),克兰费尔特氏综合征(Klinefelter's syndrome)]。组成型数量突变的发生往往是由于减数分裂中(如同源染色体配对失败)或受精中(如双精入卵——一个卵子被两个精子受精)的错误造成的。**体细胞染色体突变**(somatic chromosome mutation)影响个别的细胞及由这些细胞产生的克隆;体细胞数量突变往往反映了有丝分裂中的错误。在发育早期发生的体细胞突变可能会产生不同细胞系的嵌合体。
- 正常基因组中染色体的数目与性质用*染色体组型*(参阅)来表示,特异的畸变在通常的染色体组型附加信息来辨别。在表 4.1 中总结了这种命名法。

表 4.1 染色体畸变的细胞遗传学命名法，以人类染色体核型为例

染色体核型	含义	命名法的例子
表明染色体及染色体臂的符号		
1～22	常染色体(人类基因组)	
X,Y	性染色体	
p,q	相对的短臂与长臂	
pter,qter	短臂与长臂的末端部分	
表明突变的符号		
+	附加的染色体或染色体部分[a]	47, XY, +21
-	丢失的染色体或染色体部分[a]	46, XX, 17q-
del	缺失:末端	46, XX, del(12)(p13.3→pter)
	中间	46, XX, del(12)(p12.1→p12.3)
der	染色体衍生物(参阅)	der(12)
dup	重复	46, XY, dup(2)(q21.1→q23.1)
fra	脆性位点	46, XY, fra(Xq27)
h	附加的异染色质(即次缢痕)	46, XX, 16qh+
i	等臂染色体	46, X, i(Xq)
inv	倒位:臂间的	46, XY, inv(2)(p12.1q23.2)
	同臂内的	46, XY, inv(2)(p11.1q12.3)
marker	标记染色体[1]	47, XY, +marker
r	环状染色体	46, XY, r(19)(p13q13)
s	额外卫星物质(即随体)	46, XX, 21ps+
t	易位:交互易位	46, XX, t(7;19)(q34.2;p13.1)
	罗伯逊易位	45, XX, -14, -21, t(14q;21q)

a:“+”和“-”放在染色体标志之前，说明是额外增加或缺失的染色体(如+21)；放在染色体标志之后，则表示该染色体额外增加或缺失的不明确的片段(如17q-)。

1 译者注:原文为“未鉴别染色体”，但根据ISCN(1978)的符号和缩写术语应为“标记染色体”。

4.1 数量染色体突变

包括整套染色体的数量突变　在指定的真核细胞中整套染色体的数目被定义为**倍数**(ploidy)(表4.2)。很多多细胞的真核生物在营养阶段即指体细胞为二倍体，这是两性生殖所要求的最少的倍数。在哺乳动物与鸟类中，二倍体的组成型的改变是几乎不可耐受的，导致完全不正常的子代。这反映了常染色体与活性的性染色体之间及印痕基因间的剂量平衡(参见剂量补偿，亲代印痕)。

表 4.2 描绘倍数状态的一些术语

术语	定义
倍数(ploidy)	细胞中全套染色体的套数
二倍体(diploid)	两套染色体
一倍体(monoploid)	一套染色体
单倍体(haploid)	严格意义指在减数分裂后细胞(如配子细胞)中发现的染色体套数为正常细胞的一半。但往往作为一倍体的同义词因为大部分配子含有一套染色体(参见单倍数,一倍数)。
多倍体(polyploid)	多于两套染色体,三倍体(triploid)为 3 套,四倍体(tetraploid)为 4 套,六倍体(hexaploid)为 6 套等等。超过基因组的 10 个拷贝用数字表示,如 12 倍体(12 - ploid)。
多倍体与多线性(polyteny)	多倍体指具有超过两套同源染色体,多线性指每一染色体上有多个(等同的)染色单体,如同在果蝇唾腺中观察到的那样。
偶倍体(isoploid)与奇倍体(anisoploid)	分别指染色体套数为偶数或奇数。因为在减数分裂中不平衡的分离,奇倍体个体往往是不育的。
同源多倍性(autopoly-ploid)与异源多倍性(allopolyploidy)	同源多倍性指由内源的基因组重复引起的多倍体,而异源多倍性指由种属杂交引起的基因组重复产生的多倍体。
同倍体(homoploid)与异倍体(heterploid)	分别是具有种属典型/非典型染色体数目。

相反,在植物中单倍体与多倍体的耐受性很高,多倍体在自然中十分普遍。很明显,多倍体有增加植物大小的倾向,因而是植物育种者所需要的性状。这种性状主要是由细胞大小的增加引起的而非细胞数目的增加,表明了 DNA 含量影响细胞周期中细胞大小的关卡(参阅)。在自然界中也有单倍体与多倍体的低等动物的例子(例如,一些昆虫是单倍体,一些水蛭是多倍体),在这些动物中倍数的不正常状态较之脊椎动物中耐受性更强。多倍体动物如水蛭其细胞较之二倍体大,但由于细胞数目的减少,个体大小并不增加,这反映了发育中的调节机制(见第 6 章发育:分子机制)。

虽然具不正常的倍数的个体可能存活,但它们可能是不育的。一般,任何**奇倍体**(anisoploid,染色体套数为奇数)的种类由于减数分裂的失败都是不育的。例如,在三倍体中,三套染色体在减数分裂过程中试图配对与分离,产生严重不平衡的产物。这可应用于商业的水果作物来产生无核的品种。单倍体不能进行减数分裂只能无性繁殖。单倍体昆虫中就是这样,它们采用孤雌生殖方式,在植物中,它们的生命周期包括单倍体/二倍体世代交替。植物的单倍体品种特别有用,因为如同细菌与酵母,其培养细胞可用于突变筛选,并可根据直接显示的突变表型被选择(例如除草剂的耐受性)。具有所需性状的细胞可以通过**二倍化**(diploidization,通过人为地阻断有丝分裂诱导染色体数目加倍)得到繁殖能力,然后经过激素处理长成整株植物(见第 15 章突变与选择)。

以上讨论的多倍体形式是**同源多倍体**(autopolyploid),因为它们包含"内源"基因组的加倍。**异源多倍性**(allopolyploidy)是多倍体的一种不同形式,个体具有来自不同种属的染色体的多个拷贝。在简单的杂交失败的情况下,这种策略可将两种植物的所需性状集于一身。种间的杂交导致的杂种不育反映了减数分裂的失败,即不同种属的染色体不能形成合适的配对。然而如果杂合的基因组被诱导加倍,那么每一个染色体都有一同源的染色体可配对,杂交种属就成为有繁殖能力的了。在同源多倍体中染色体之间种属特异

的配对被称为**同源联会**(autosynapsis),而种属之间的配对被称为**异源联会**(allosynapsis)。来自不同种属的部分同源的染色体被称为**部分同源染色体**(homoeologous chromosome)(参考同源染色体)。

表 4.3 用于描绘核型中个别染色体的数目的术语

术语	定义
整倍体(euploidy);非整倍体(aneuploidy)	具有整套染色体;个别染色体丢失或获得。
染色体不平衡(chromosome imbalance)	产生多重基因剂量效果的染色体的丢失与获得。
超倍体(hyperploidy);亚倍体(hypoploidy)	具有超过正常数目的染色体;具有少于正常数目的染色体。
假二倍体(pseudodiploidy)	具有正常的染色体数目但核型不正常,由于同时发生了不同染色体的单体性与三体性。
正体(eusomy);非正体(aneusomy)	特定染色体具有正确的拷贝数;特定染色体具有不正确的拷贝数。零体(nullisomy)为无拷贝,单体(monosomy)为一拷贝,双体(disomy)为二拷贝,多体(polysomy)为两拷贝以上,如三体(trisomy)为三拷贝,四体(tetrasomy)为四拷贝,等等。

表 4.4 人类出生存活时观察到的非整数倍

非整数倍	发生率与临床表型
常染色体非整数倍	
21 三体(唐氏综合征,Down's syndrome)	在出生存活婴儿中发生率为 1:700,临床表型表现为智力与生长障碍,特征性脸部特征(眼上吊,鼻小,张口伸舌,耳位低),部分并趾具有类人猿的掌纹。再显风险率很低虽然它随母体年龄增加而增大。如果三体是由于易位引起的(参见*易位唐氏综合征*),则再显风险率较高。21 三体患者的寿命预期比正常人低,很多在儿童时期死亡。
18 三体(爱德华综合征,Edward's syndrome)	在出生存活婴儿中发生率为 1:3~6000[1],临床表型为长头,耳位低且发育异常,罗圈腿,紧握拳,心血管与肾脏功能紊乱,产期死亡。
13 三体,偶尔 14 或 15 三体(帕韬氏综合征,Patau's syndrome)	在出生存活婴儿中发生率为 1:5~10000[1],临床表型为耳位低且发育异常。产期死亡。
性染色体非整数倍	
单体 X(特纳氏综合征,Turner's syndrome)	在女性出生存活婴儿中发生率为 1:10000。临床表型为身材矮,蹼颈、支撑角宽(肘外翻),乳头距离远,基本无月经,且不育
三体 X(XXX 综合征)	在女性出生存活婴儿中发生率为 1:1000。临床表型包括一级与二级性征发育失常,轻度智能障碍,但很多个体是可育与正常的。
多体 X 和 Y(克莱恩费尔特综合征,Klinefelter syndrome)	主要与核型 47, XXY 有关,虽然已观察到更多包括多个 X 与 Y 染色体的非整数倍体。在男性出生存活婴儿中发生率为 1:1000,临床表型可能包括不育,男性乳房发育(在男性中胸脯发育),睾丸萎缩,阴茎畸形,精子无功能,身材较高,轻度智能障碍。X 染色体数目增加,表型趋向严重。
二体 Y("超雄"综合征,"supermale" syndrome)	虽然观察到稀少的 48,XXYY,但主要与核型 47,XYY 有关。在男性出生存活婴儿中发生率为 1:1000。临床表型包括身材较高。核型与反社会及暴力行为倾向有不明显的联系,这种联系仍未定论。个体通常可育。

1) 这两处数字按原文可能有误,似应为 1/6000~3/6000 以及 1/10000~5/10000。——译者注

包括个别染色体的数量畸变　一正常细胞，**整倍体**(euploid)应该包含整套染色体，如发生个别染色体缺失或增加，就被称为**非整倍体**(aneuploid)(表4.3)。非整倍体在植物或动物中都是有害的，而在动物中后果尤为严重。染色体的缺失或获得被称为**染色体不平衡**(chromosome imbalance)，这样会发生多个基因的缺失或获得，具多重剂量效应。在人类中这种效应往往是致命的，在可能的22种常染色体三体中，只有三种能在出生时存活，只有一种可长期存活(表4.4)。因为剂量补偿机制(参见)可以抑制性染色体数目天然变化的后果，性染色体的非整倍体耐受性较高。而由于X染色体失活(参阅)X染色体的单体性与多体性产生相对温和的表型。在特纳综合征(Turner's syndrome)与XXX综合征中的失常可能反映了逃脱了失活的基因的剂量效应，并且效应发生在发育早期X失活之前。Y染色体的二体性几乎没有表型效果，与这种核型相联系的行为异常仍有争议。

数量染色体畸变的起因　倍数的改变可能由细胞周期的失误引起，或是由DNA复制遗漏/额外的染色体分离(使倍数减半)引起，或由染色体分离遗漏/额外的DNA复制(使倍数加倍)引起。倍数的改变也可能由受精过程引起，如在哺乳动物中，三倍体可能是由于精子或卵子的二倍性造成的，更常见的是发生二精入卵(由两个精子受精)。这种由不正常的受精过程导致的发育不正常被称为**囊状的或水泡状脂块**[hydatidiform或vesicular mole(参见亲代印痕)]。

非整数倍反映了细胞周期更为局部的紊乱并可以几种方式发生，最常见的被称为**不接合**(nonconjunction，在减数分裂的前期同源染色体对未能相互接合，参见减数分裂)与**不分离**(nondisjunction，配对正常，但同源染色体对分离失败)。**初级不分离**(primary nondisjunction)发生在第一次减数分裂中，导致某一对同源染色体移向一极。**次级不分离**(secondary nondisjunction)发生在第二次减数分裂中，导致了一对姐妹染色单体一起移向同一极(图4.1)。这往往是由于着丝粒分离的失败。在有丝分裂中也会发生类似的情况(**有丝分裂不分离**，mitotic nondisjunction)，导致体细胞的非整倍体。如果着丝粒过早分离，在第一次减数分裂中也可能发生部分的不分离，将一对染色单体中的一个与另一同源染色体的一对染色单体分到一起。几种结构突变增加了不分离的机会(参见环状染色体，并连染色体，罗伯逊易位)。

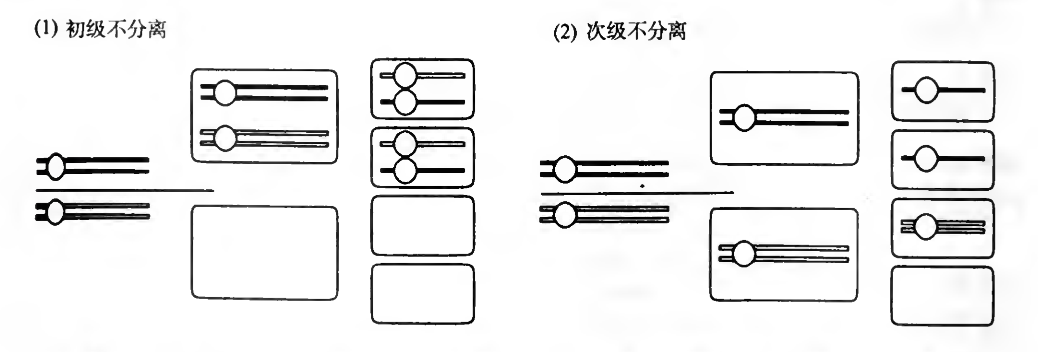


图4.1　在配子中由于不分离现象引起的非整倍体。(1)在第一次减数分裂中同源染色体未能分离(初级不分离)(2)在第二次减数分裂中姐妹染色单体分离失败(次级不分离)。

引发非整倍体的另外一些途径包括**染色体获得**(chromosome gain,单个染色体的额外复制的结果)及**染色体丢失**(chromosome lose),后者发生于某一染色体分离过慢而未包括在子代的核中(**后期落后**,anaphase lag),或染色体未能在第一位置与纺锤体相接。

4.2 染色体结构突变

断裂点 染色体结构突变是由染色体断裂的错误修复或同源非等位位点之间的重组造成的。发生断裂或重组的位点被称为**染色体断裂点**(breakpoint),结构重排的染色体被称为**染色体衍生物**(chromosome derivative)。如果衍生物包含不同染色体的部分,它们根据着丝粒的来源命名。那么,如果人类 12 和 17 号染色体断裂并且断裂片断交换,包含 12 号染色体着丝粒的染色体被称为 12 号染色体衍生物。上面已讨论过,结构突变可能是平衡的,也可能是不平衡的。平衡与不平衡的突变型都可用遗传学方法根据它们的改变的连锁图谱来辨别,并且在杂合子中,可用细胞学方法通过减数分裂中同源染色体的异常配对图式来辨别。动物与植物对结构突变的耐受性不同。即使是严重不平衡的动物配子也仍具有功能,并可参与受精过程产生不平衡的合子。植物中如突变发生在胚珠中情况也一样,但发生染色体结构畸变的花粉粒却容易退化,并不发生受精。

不平衡结构突变 **不平衡突变**(unbalanced mutation)包括遗传物质的丢失或获得,表现为部分的单体性或三体性 - 即由于一些连续基因的丢失或获得,它们表现出多剂量效应。包括 DNA 丢失的结构突变被称为**缺失**(deletion)或**缺乏**(deficiency),而包括 DNA 获得的结构突变被称为**插入**(insertion)或**重复**(duplication);如果它们涉及遗传物质的量是可见的,它们就被认为是染色体突变。大的缺失可以通过在杂合子中**缺失环**(deletion loop)的出现来辨别,这是出现在减数分裂中由于没有同源配对区域而突出于配对染色体的一个结构。插入也有类似的表现,而串联重复可以产生更加复杂的结构,重复结构可能会突出形成一个环,也有可能染色体内部的重复序列的两个拷贝相互配对而使正常染色体的同源区域环状突出(另参见*基因扩增*)。

在人类染色体中最小的可见缺失与重复都会影响几兆碱基的 DNA,因此覆盖很多基因位点。它们往往引起**连续基因综合征**(contiguous gene syndrome)(表 4.5)。这些突变通常没有精确定义,因为在细胞学水平无法确切了解它们的边界。连续基因综合征可以通过包括一系列相互覆盖的临床症状的可变表型来辨别。不同病人中个别症状的对应关系使我们可以构建一基因图谱,被称为**发病图谱**(morbid map)。如果在没有可见突变的情况下发生同样的表型,就可在光学显微镜的分辨率之下推测缺失或重复的大小(**微小缺失与微小重复**,microdeletions and microduplication),或这种表型可能由*点突变*(参阅)引起。

当着丝粒在错误的平面分裂时,同时发生整个臂的缺失与重复,产生具有两个长臂没有短臂的染色体及相反的情况。这种具中间着丝粒的染色体被称为**等臂染色体**(isochromosome),它引起一臂的单体性及另一臂的三体性,从而产生相应的基因剂量效应。

平衡结构突变 **平衡突变**(balanced mutation)又被称为**染色体重排**(chromosome rearrangement),它包括 DNA 的重排而没有遗传物质的净获得或失去。带有这种重排的个体

表 4.5 与不平衡结构染色体畸变(缺失与重复)有关的人类综合征

综合征	核型与临床症状
猫叫综合征(*Cri du* chat syndrome)	del(5)(p14-pter)。小头,智能障碍,特征性哭声似猫叫
贝-魏综合征(Beckwith-Wiedemann syndrome)	dup(11p15)。胎儿过度生长导致巨人症,舌头肥大,儿童时期出现侧位肿大
WAGR 综合征(WAGR syndrome)	del(11p13)。根据主要的症状命名:维尔姆斯瘤,无虹膜,两性生殖器官及智能障碍
猫眼综合征(Cat eye syndrome)	+ iso(22)(pter→q11)。眼缺陷,肛门闭锁,心血管缺陷和智能障碍
Prader-Will 综合征	父母染色体 del(15)(q11 - q13)。身材矮小肥胖,生殖器官小,行为困难,智能障碍
Angelmam 综合征	母系染色体 del(15)(q11 - q13)。智能障碍,运动失调,不适宜地发笑

某些病症,如 WAGR,是真正的连续基因综合征,一系列相互覆盖的临床症状反映了一簇连续基因的缺失。另外一些,如贝 - 魏综合征,主要与某一单个基因的缺陷有关(在这个例子中是 IGF2 基因的活性过高;参见*亲代印痕*)。

往往具有正常的表型,但杂合子可能是半不育的。因为:①在减数分裂中重排染色体表现不正常,往往分离产生不平衡的配子;② 重排内的交换产生不平衡的重组产物,倒位不能存活。由于空间位阻(steric hindrance)在这些重排的断裂点重组受抑制(参见*重组冷点*)。这些原则应用于果蝇来抑制交换,使特定染色体可以在不重组的情况下传递,以获得特定的果蝇株型。具有复杂的多个倒位的染色体可抑制交换因此每一个染色体可作为完整的单位遗传。这种重组抑制因子被称为**平衡器染色体**(balancer chromosome)。

单个染色体的平衡结构突变　　倒位(inversion)指包括某一单个染色体中 DNA 片断的翻转的染色体重排过程。它们可能由染色体断裂的错误修复或反向重复序列之间的染色体内重组引起。除非①断裂点打断了一个基因;②重排导致不利的*位置效应*(参阅),倒位不产生表型,但倒位序列可以通过遗传学方法用改变连锁图谱来辨别,或在杂合子中通过观察减数分裂中**倒位环**(inversion loop)的形成可辨别大的倒位,倒位环指染色体沿轴的方向弯曲,以使倒位位点能插入配对染色体相应区域。

因为倒位只涉及单个染色体,本质上它们不会影响减数分裂的分离(参见*转座*)。然而,由于严重不平衡的交换产物的产生,它们抑制重组回复。交换产物的本质取决于倒位区是否含盖着丝粒(图 4.2)。如着丝粒包括在倒位中(**臂间倒位**,pericentric inversion),就会产生包含缺失与重复的重组产物。然而,如果着丝粒不在倒位发生范围内(**臂内倒位**,paracentric inversion)就存在潜在的更严重的问题,重组可能会把两个着丝粒连在一起形成一个有两个着丝粒的染色体和一个无着丝粒断片。后者由于无法附着于纺锤体上往往会丢失。有两个着丝粒的片断会延伸在两个核之间,或者由于后期落后而被排斥,或者被随机打断产生一断裂端,可能会起始另一个*断裂 - 融合 - 桥循环*(参阅)。

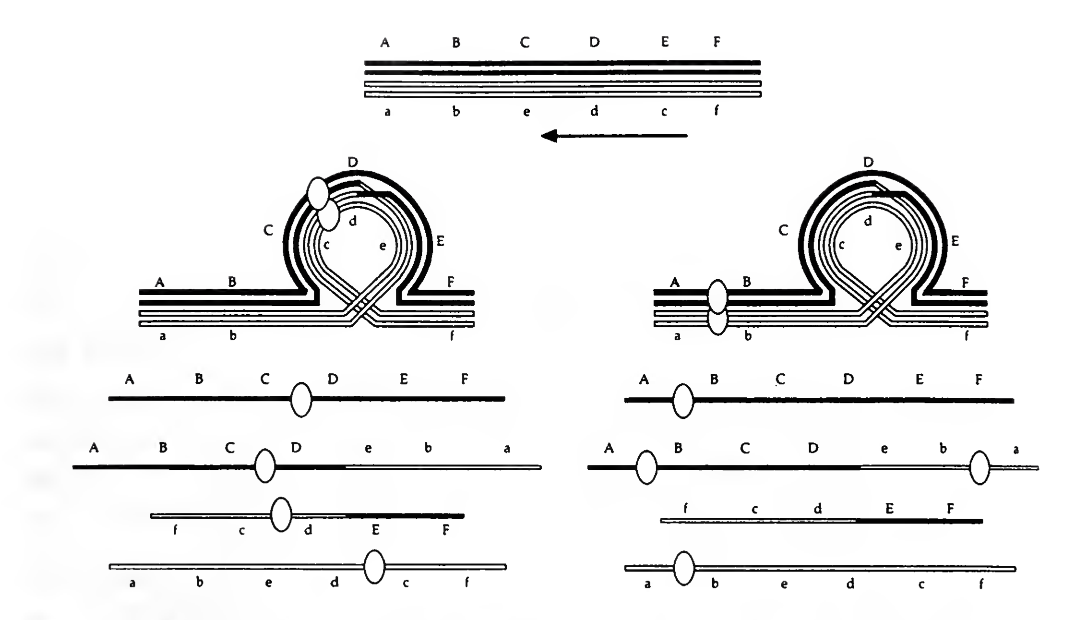


图4.2 在倒位杂合子中交换的效果。倒位使一染色体上位点C,D,E方向颠倒,在减数分裂时形成一倒位环。在两侧倒位中(左),着丝粒包括在倒位中,交换产生包含缺失和重复的不平衡产物。在臂内倒位中(右),着丝粒在倒位范围之外,倒位环内发生交换产生一无着丝粒断片(将丢失)和两着丝粒的桥。

第二类分子内的结构畸变的情况为染色体两端的端粒都丢失,两个断裂的末端融合到一起。这种环状染色体复制后连成串,如果不能分开的话,不分离是不可避免的。

多个染色体的平衡结构突变 一条染色体的一部分转移到另一染色体上,这种结构畸变被称为易位(translocations)。易位可能是交互的也可能是非交互的。当两条染色体断裂,远端片断互相交换并重新连接,这种情况是**相互易位**(reciprocal translocation)。**非相互易位**(nonreciprocal translocation)指物质的单向转移。一条染色体片断插入另一染色体的内部位点被称为**内易位**(internal translocation)。同样的,遗传物质加到另一染色体的末端被称为**末端易位**(terminal translocation)。**跳跃易位**(jumping translocation)是一种特别的末端易位,同样的染色体片断连续地跳跃到不同染色体,在不同细胞中出现在不同末端位置。非相互易位的一种特别形式被称为**罗伯逊易位**(Robertsonian)或**整臂易位**(whole arm translocation),两条近端着丝粒染色体(参阅)在着丝粒或着丝粒附近融合产生一可作为单个分离单位分离的**复合染色体**① (compound chromosome)。

如同倒位一样,易位是平衡的。但与倒位不同的是,易位涉及一条以上的染色体。因此,在杂合子中减数分裂时的配对可能会包括两对以上染色体。因而易位杂合子往往不育并不仅仅是由于交换的效应,也因为分离过程本身不正常,会产生不平衡的配子(图4.3)。由于罗伯逊易位涉及整条染色体,不平衡的分离产生单体与三体配子。这可由唐氏综合征的发病率反映出来,当由自发不分离引起时其再显率很低,但如受包括21号染色

---

① **复合染色体**指融合在一起的两条完整的染色体。罗伯逊易位包括两条近端着丝粒染色体(参阅)的融合,部分短臂的丢失对表型没有明显的影响。两条长臂融合产生近中着丝粒染色体(参阅)包含两条原来的近端着丝粒染色体的长臂及一通常的着丝粒(因此整臂易位)。也有可能两条近中着丝粒染色体,如X-染色体,连接在一起。这样的染色体是具双着丝粒的,被称为**并联染色体**(attached chromoscme)。并联-X染色体可用于证明遗传的染色体理论。

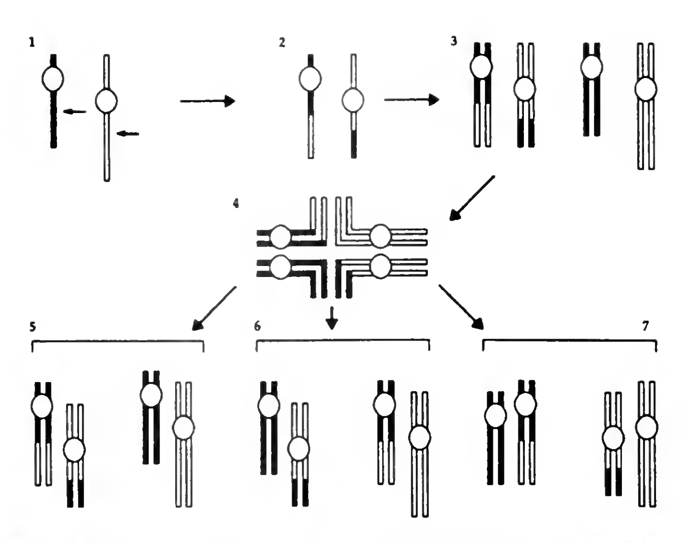


图 4.3 在减数分裂中杂合子的交互易位行为。(1) 两条染色体发生了双链断裂(箭头所示),然后(2) 末端连接修复引起了一交互的易位(或者易位是由于每一染色体中位点间的重组导致的)。(3) 在减数分裂中,易位体试图与它们正常的同源序列排列在一起,通常产生一四染色体结构被称为**四价染色体**(tetravalent)。(4) 四价染色体可能以多种方式分离。在**相间分离**(alternate segregation)中(5),两条正常的染色体分到一极,而两条发生易位的染色体分到一极,产生平衡的配子。在 **1 类相邻分离**(adjacent - 1 segregation)中(6),每一正常的染色体与它的非同源的易位染色体分到一起,产生不平衡的配子。在 **2 类相邻分离**(adjacent - 2 segregation)中(7),每一正常的染色体与它的同源的易位染色体分到一起,产生不平衡的配子(2 类相邻分离极少发生)。在罗伯逊易位中(图中没有显示),复合染色体试图与其同源序列配对,形成一三染色体结构被称为**三价染色体**(trivalent)。在相间分离中,复合染色体分离到一极,另两条正常的染色体分配到另一极,产生平衡的配子。在相邻分离,复合染色体与正常的同源染色体的一条分离到一极,剩下的另一条正常的染色体分配到另一极,产生配子不平衡且是非整数倍的。

体的罗伯逊易位引发,再显率要高很多。在最特殊的例子中,罗伯逊易位可将 21 号染色体的两个同源拷贝融合在一起。在这种情况下,每一次减数分裂都会产生 21 号染色体不平衡的配子,所有的存活后代都患有唐氏综合征(即,再显风险率为 100%)。

脆性位点　染色体上不能染色的空隙被辨明是严厉生长条件诱导的次级缢痕,人们认为这是染色体断裂的热点,被称为**脆性位点**(fragile site)。这些位点根据诱导它们产生的生长条件来分类。在人类染色体中,已辨别出出现在大部分人体内的超过 50 种**共同脆性位点**(common fragile site)。还描绘了出现在人群中少数个体中的 20 种**稀有脆性位点**(rare fragile site)。一般说来,脆性位点没有表型效应。一个例外是位于 Xq27 的叶酸敏感的稀有脆性位点,对应地在男性携带者中出现智力障碍,大睾丸以及特征性的面貌;在一些女性携带者中出现智力障碍[**马丁 - 贝尔综合征**(Martin-Bell syndrome)或称**脆性 X 综合征**(fragile X syndrome)]。这种综合征的病因是基因 *FMR1* 中的多态性的三联重复突变,这就解释了观察到的复杂的遗传图式(参见三联重复综合征)。虽然影响 DNA 复制与修复的条件可以诱导产生脆性位点,但使这个座位具有脆性的位点性质本质上与这种突变没有关系。

结构突变的起因　缺失、重复、倒位或易位都可能由不正确修复的 DNA 或同源的畸变形式或位点特异重组引起(参见异常重组)。串联重复 DNA 的不等交换可能产生缺失与

复制，同时分散的重复序列如不同染色体上的转座因子之间的重组会导致易位。易位也可能是有非法的 V-D-J 重组(参阅)引起的，这种畸变往往是淋巴样失常的根源(见第 18 章癌基因与癌)。

染色体的断裂可能自然发生，也可由被称为**断裂剂**(clastogen)的化学或物理因素诱导发生。在具有遗传性弱点时，如负责 DNA 重组与修复的基因的突变，发生断裂的机会上升。染色体断裂的频率与**姐妹染色单体交换**(sister chromatid exchange，**SCE**，染色体中两姐妹染色单体之间的交换)的频率有关，因为同源重组会被双链断裂所促进(见第 25 章重组)。SCE 的频率可通过姐妹染色单体的不同染色来定量(参见复制带型)。在几种染**色体不稳定综合征**(chromosome instability syndrom)中观察到 SCE 的频率上升：Bloom 氏综合征，Fanconi 氏综合征和共济失调毛细血管扩张，它们是由 DNA 修复基因或 DNA 损伤反应基因的突变所至(参见修复缺陷综合征)。染色体 DNA 中断裂位点的分布并不是随机的。虽然已知不同的断裂剂具有不同的断裂热点，染色质区域中的热点簇主要位于明 G 带(参见染色体带型)，暗示了敏感位点与染色质结构有关(见第 3 章染色质)。

个体中的混合核型　　上面几节讨论了**组成型染色体突变**(consititutional chromosome mutation)，即影响体内的每个细胞。第二类染色体畸变被称为**体细胞染色体突变**或**获得性染色体突变**(somatic or acquired chromosome mutation)，这种畸变在发育中发生，导致了同一核型中不同核型的混合(表 4.6)(参见体细胞突变，癌症)。这样的一个体被称为**嵌合体**(mosaic)。同样的混合核型也可能在两个独立发生克隆的混合情况下发生。这样的个体被称为**异源嵌合体**(chimera)。在两种情况都是同一生物体中分布着不同的细胞，但只有在嵌合体中，两类细胞才来自同一祖先。异源嵌合现象天然发生的频率很低：异卵双生的融合导致**双精子嵌合体**(dispermic chimeras)的产生，源于同对双生细胞的一双生体的克隆化导致**血细胞嵌合体**(blood chimeras)的产生(另参见转基因小鼠，ES 细胞)。

表 4.6　各方面等同的细胞表现的不同表型有关的术语及它们的内在起因

术语	定义
花斑(variegation)	在一组织中，出现表现不同表型的部分(例如一些植物中的白/绿花斑)。这可能并不反应基因型的内在差别。
混倍性(同体异数)[mixoploidy (aneusomaty)]	携带具有不同核型的体细胞的个体。混倍性可以是嵌合体也可以是异源嵌合体(见下文)。**真体**(eusomaty)指细胞具有相同核型的个体。
嵌合体(mosaic)	携带来源于同一祖先具有不同基因型(或核型)的体细胞的个体，如来源于突变或体细胞重组。
异源嵌合体(chimera)	携带来源不同具有不同基因型(或核型)的体细胞的个体。异源嵌合体往往来源于胚胎融合或克隆化(参见转基因小鼠，ES 细胞)。
非自主表型(allopheny)	由于另一组织中的嵌合导致的一组织表现出花斑的现象，例花斑组织的差异不是由于内在的差异，而是由于具有不同基因型的其他细胞的诱导性的相互作用。获得非自主表型的动物可测试一突变的细胞自主性。非自主表型的果蝇可以通过 X 射线诱导体细胞重组产生，例如为了研究眼发育中的相互作用(参见 sevenless 信号途径)。

混合核型效应的最重要的决定因素是各细胞株的克隆程度及涉及的基因自主性。一

有害的核型如果只限于组织的一小部分或这种缺陷是细胞非自主的(即它可被其他细胞的产物所补偿)情况下可能不会有整体的效应。反之,一核型如果是广泛存在或它影响一特异的器官,或这缺陷是细胞自主的从而导致广泛的细胞死亡,它有可能是致命的。在动物中,来自雌性和雄性的细胞的嵌合体或异源嵌合体(**雄雌嵌合体**,gynandromorph)可以象任一性别那样正常发育表现为间性(intersex,具有两种或中间性征的个体)或表现为**雄雌同体**(hermaphrodite,既可以产生雄性配子又可以产生雌性配子的个体),这依赖于每一细胞株的克隆程度。

嵌合体这个术语常常被广泛地用来描绘任何各方面等同的细胞表现的不同表型的情况。然而严格说来嵌合现象指基因型的混合而不包括由不同的外显率与表现度引起的不同表型,如由环境效应或后生调节引起的表型差异(例如*位置花斑效应*(参阅)或不同活化与失活的 X 染色体的遗传)。当未监测到基因型的内在差别时,应该使用表型**花斑**(variegation)这个术语。

(徐　沁 译)

第5章
染色体结构与功能

基本概念与定义

- **染色体**(chromosome)是一独立的携带必需遗传信息的DNA分子,并包括决定其结构的有关蛋白质。这个术语的范围可包括细菌的核酸,细胞器基因组和病毒基因组,同样包括真核核内染色体。在这一章中只讨论最后一种情况。
- 最初认识到基因由染色体携带是基于染色体与孟德尔基因行为的平行关系(成对出现,均等分离,独立的类型)。这是**遗传的染色体理论**(chromosome theory of inheritance)。
- 在二倍体真核生物中有三个水平的对偶性。体内出现同源染色体对,并且当单个染色体在减数分裂中变得可见时,可以看到它们包括一对**姐妹染色单体**(chromatids),由一共同的着丝粒连接,每一姐妹染色单体包括两条双螺旋的DNA链。尽管出现了八条DNA链,细胞仍然是二倍体。事实上,每一双螺旋中只有一条链携带有(转录)信息,并且两个姐妹染色单体的出现代表了信息的加倍以备分离进入子细胞-细胞产生时只带有一份拷贝(当分离时姐妹染色单体成为一染色体)。另外,每一有丝分裂染色体的姐妹染色单体是等同的,而同源染色体不等同。因此在整个细胞周期中,每一位点只有两个等位基因。
- 在减数分裂中可辨别出两类染色体:**常染色体**(autosome)具有**同形**(homomorphic,结构完全一样)的配对染色体因此形成同源染色体对。**异染色体**(heterosomes or allosome)具有**异形**(heteromorphic,结构不同)的配对染色体并且只有部分同源区能相互配对。两条异形染色体的相互配对部分被称为**假常染色体区域**(pseudoautosomal region),因为这个区域中的基因表现出常染色体基因同样的遗传方式。异染色体往往是**性染色体**(sex-chromosome)(参见谱系分析、性连锁遗传、性别决定)。
- 染色体的作用是提供一框架,使基因组的线性片断能有效地复制与分离。这两个过程中任一的失败都会导致子细胞中的染色体不平衡。三个特异的顺式作用位点是稳定的染色体维持所必需的,它们是复制原点(参阅)、着丝粒、端粒。

5.1 正常的染色体——一般形态

正常染色体的细胞遗传学特征 在细胞周期的大部分中,真核染色体以松散堆积的染

色质形式存在，并且无法从细胞核中辨别出来。它们在有丝分裂（或减数分裂）的开始成为可见，这时染色质压缩后形成分立的结构，当用适当的染料处理核时可以被染色（文框5.1）。细胞分裂中期的染色体的形态特征在图 5.1 中显示，这使我们可以识别各个染色体并辨别畸变（见第 4 章染色体突变）。

表 5.1 染色体分带技术的选择

分带技术	方法与应用
C 带型	一基于吉姆萨方法的技术，包括与氢氧化钡的保温，可辨别异染色质（参阅）区域，因此可用来确定着丝粒的位置。
C_d 带型	辨别动粒（参阅）的技术。
D 带型	辨别 DNase I 敏感区域的技术，一般对应于具有潜在转录活性的开放染色质区域。
G 带型	最常用的染色体分带技术，高重复性的带型是国际标准人类与小鼠细胞遗传学图谱的基础。染色体用胰蛋白酶部分水解，与**吉姆萨氏染料**（giemsa's stain，溶于甲醇中甲基蓝，曙红及其他染料的混合物）保温。
G12 带	在高 pH 下的 G 带技术，对人类与小鼠的染色体染色不同，使它们可以区别开，例如在体细胞杂交中。
N 带型	特异染色核仁组织区域的技术，如硝酸银染色。
Q 带型	用荧光染料奎纳克林的带型技术。它产生与 G 带型相同的图谱，但还使异染色质区域发荧光，如辨别 Y 染色体。
R 带型	一基于吉姆萨方法的技术，包括一热变性步骤获得与通常的 G 带型方法相反的带型图式。这技术非常有用，如鉴别末端缺失等。
复制带型	这个技术包括将溴脱氧尿苷插入复制 DNA。在 S 相的一短的脉冲是复制时间可被监测。如果延长时间，产生与吉姆萨染色相似的图式，说明 G 带对应于复制时间区带（参阅）。插入后经过两轮复制可以分辨姐妹染色单体由于一条染色体两条链都有溴脱氧尿苷的插入，另一只有一条链被插入。这样姐妹染色单体在吉姆萨染料或在荧光染料 Hoechst 33258 的存在下就会有不同的染色情况。这个技术被称为**花斑染色**（harlequin staining）可用于姐妹染色单体交换（参阅）的监测。
T 带型	R 分带技术的变化。只有端粒 DNA 可被染色，T 带是基因富集的区带（参见等容，转录图谱）。

在很多情况下，这些方法是经验性的，即它们的工作原理并未被了解，值得注意的是这些不同的技术产生同样的相似的带型，强烈暗示了使基因组成为分立的等容的结构组织。

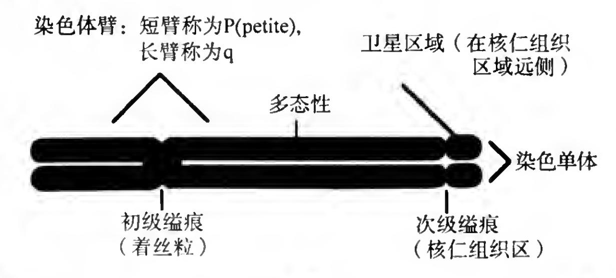


图 5.1 正常中期染色体的形态特征。**初级缢痕**（primary constriction）指明了着丝粒的位置染色较浅，连接所有四条**染色体臂**（arm）。**次级缢痕**（secondary constriction）染色较浅，往往代表了**核仁组织区域**（nucleolar organizer region，NOR），是串联重复 rRNA 基因的位置。在人类染色体 1，9 和 16 中发现了中间的次级缢痕，而在染色体 13，14，15，21 和 22 中发现了远端次级缢痕。NORs 的远端端粒染色体片断被称为卫星区域因为它们与染色体的主体显得被分离。由于核仁组织区形成一共同的**核仁**（nucleolus），不同染色体的**卫星区域**（satellite region）往往连接在一起（**卫星连接**，satellite association）。其他的次级缢痕可由在不正常的条件下细胞的生长所诱导，被称为脆性位点（参见）。**染色体多态性**（chromosome polymorphism）或称**异形性**（heteromophism）是在一个群体中变化的染色体的某些遗传形态特征，它们没有表型效应。例子包括可变异染色质（如图所示），倒位（参见）和稀有脆性位点，这些可用作细胞遗传学标记。

染色体带型　在图 5.1 中显示描绘的染色体的特征基于与 DNA 结合染料如富尔根(feulgen)的**均相染色**(homogeneous staining)。哺乳类的染色体也可使用一系列破坏性技术以异相染色方式被染色,会显现明暗相间的高重复性与特异的横带图式(表 5.1)。这种**染色体分带**(chromosome banding)方法是具有同样的大致形态的染色体可被分辨,并为基因作图提供了一细胞遗传学的基础。它们在表征染色体畸变时特别有用。在细胞分裂中期的染色体中只能观察到低分辨率,但在较不紧密的前中期染色体中每一中期带型都可细分为许多亚带。**人类细胞遗传学命名法**(the International System for Human Cytogenetic Nomenclature,**ISCN**)就是基于早、晚前中期与中期染色体的带型与亚带型建立的(分别具有 400、550 与 850bp 的分辨率)。每一带用它所在的染色体与染色体臂(例,1p,2q)和从着丝粒数起的带数来表征(*cen*)。分辨率增加用更多的数字来表示(例如 1p3→1p34→1p34.1)。染色体带型图式反映了哺乳类基因组的结构与功能的组织(参见等容),例如,明的吉姆萨带对应于具有一般转录活性、早期复制与低重复 DNA 含量、DNA 酶敏感的区域。由于每一条带对应于超过 10Mb 的 DNA,这种是大致的划分(见第 3 章染色质,第 12 章基因组与作图)。低等的真核染色体用适用于哺乳动物染色体的染色技术不能染色,但双翅目昆虫的多线染色体显示天然的高分辨率的带型(见下文)。

5.2　特异染色体结构

A 和 B 染色体　在大多数真核生物中,正常个体的每一个细胞中都携带有确定不变的一套染色体,可以用于种属的判断(参见核型;参考双微染色体,基因扩增)。然而有些种属携带有额外的染色体,往往表现为全部由异染色质组成,没有表型效应,在指定群体中甚至在同一个体的不同细胞中往往都有变化。这些结构有不同的名称,如**附属染色体**(accessory chromosomes),**卫星染色体**(satelite chromosomes),**超数染色体**(supernumerary chromosomes),或者在植物中,被称为 **B 染色体**(B-chromosome)(与不变和必需的 **A 染色体**(A-chromosome)相区别)。它们一般不参与减数分裂而是随机分离。它们可以被认为是巨大的线性质粒。

多线染色体　在果蝇与其他双翅目昆虫中,特定幼虫分泌组织(如唾液腺)的细胞中含有由超过 1000 个染色单体组成的**巨大染色体**(giant chromosome)。它们也可被称为**多线染色体**(polytene chromosome)或根据它们的发现者命名为**巴尔比亚尼染色体**(Balbiani chromosome),这种染色体是由于作为一种基因扩增的策略在不发生有丝分裂情况下多轮复制而造成的。多线染色体不仅比正常染色体更粗,也比正常染色体更长,由于多个染色单体的聚集使正常的染色质结构不可能形成。

多线染色体的一显著的特征就是高度重复性的横向深色的**带**(band)(或称**染色粒**,chromomere)和浅色的**间带**(interband)。它们被认为反映了染色质密度上的区域差异并且功能性地对应于不同的染色质区域(参阅)。这种带型的分辨率甚至较之哺乳类染色体的前中期的带型更为精细(各多线带长度 1 到 100kbp DNA 的范围内),但可以相同的方式应用,来构建细胞遗传学图谱与表征染色体突变。

多线染色体的第二个特征是**染色体膨大**(chromosome puff)和**巴尔比亚尼环**(Balbiani ring)。它们分别是染色体上相对较小或较大的膨胀部分,对应于转录活性区域,可能反映了染色质结构的区域膨胀。特定的环境刺激可诱导发生重复性的膨大或环的图式(例用蜕皮激素处理),这样可以通过直接观察确定激活基因的位置。

灯刷染色体 在双翅目的卵母细胞中,细胞周期在第一次减数分裂的双线期(参阅)被抑制,形成单个的核仁以满足特大细胞的需要。减数分裂的二价体的四个染色单体被交叉连在一起,可以看到,从深色的堆积的染色粒(同一术语用来描绘灯刷染色体与多线染色体中深色堆积的可能是非活化的染色质部分,也用在去紧缩的哺乳类染色体上)突出长的不卷曲的转录活化染色质的**外侧环**(lateral loops)。这种环成对出现,发生于每一染色单体上,并且它们被一核糖核蛋白的环所环绕。这些以及逐渐去紧缩的环结构被认为反映了染色体连续与巨大的转录活性。整个细胞共有 10—15 000 在活动的环,但大部分 DNA 仍以染色粒的形式紧缩。如同多线带与间带,每个环被认为对应于一染色质区域(参阅)。

5.3 染色体结构的分子方面

染色体功能所要求的分子结构 染色体的大体形态对它如何在细胞中起作用提示很少。首先,染色体的功能是提供了一个框架,使基因组的每一线状片断可以有效地复制与分离,这两个过程中任一的失败都会导致子细胞中染色体的失衡(见第 4 章染色体突变)。由于每一真核染色体带有基因的不同排列,染色体上携带的信息不会影响它的功能。稳定维持的染色体要求有三个顺式作用位点:复制起点(参阅),着丝粒和端粒。复制原点对染色体的复制是必要的,因为它提供了复制起始蛋白质的结合位点(见第 26 章复制)。**着丝粒**(centromere)对分离是必要的,因为它提供了一动粒组建的位点,促进微管的附着。**端粒**(telomeres)对染色体的稳定是必需的,因为它们在 DNA 复制后完成染色体末端形成,防止与其他染色体错误的末端连接。携带有任意 DNA 的人工染色体(参阅)只要具有这三种元件,就可以与内源基因组共同稳定保持。

着丝粒的分子性质 在酿酒酵母中,遗传学上着丝粒根据它们在质粒中的有丝分裂的稳定性来确定(这被称为 **CEN 功能**,CEN function)。用染色体步移(参阅)的方法克隆了几个着丝粒,它们显示在功能上可相互代替。序列比较辨明了三个保守元件,被称为 CDI,CDII,CDIII。CDI 具有一短的共有序列主要在减数分裂中起作用。CDII 是一约 100bp 的富含 AT 的序列,其作用是不清楚的。这两种元件中的突变影响有丝分裂的过程但不会消除它,CDIII 是一个高度保守的 26bp 的元件表现出二重对称性,它似乎对着丝粒的活性是必需的,因为在对称中心的点突变能消除着丝粒功能,导致不稳定的分离。一个被称为 Cbf-III 的多蛋白复合物与这元件结合并表现微管马达活性。Cbf-III 可能代表了微管附着与分裂后期分离启动的位点。

粟酒酵母及更高级的真核生物中,着丝粒延伸约几十 kbp,而相对的较小的酿酒酵母的着丝粒为 125bp,这反映了纺锤体附着的性质。在酿酒酵母中,一单个纺锤体纤维附着于每一染色体,而在粟酒酵母及哺乳动物中,涉及数目众多的纤维,着丝粒中包含重复 DNA。在粟酒酵母中着丝粒表现出二重对称性,而在哺乳动物中,着丝粒包含大量的卫星 DNA(参

阅),在灵长类动物中包含170bp局部微扰的串联重复。某些蛋白质特异地作用于哺乳类着丝粒,其中之一在体外能与卫星DNA结合,并可能代表了着丝粒形成位点。

端粒的分子性质 线性染色体的属性代表了真核细胞的两个问题。首先,由于DNA聚合酶(参阅)的性质,每一链的5′端在复制中不能完成(见第26章复制)。第二,由于DNA修复的末端连接酶的丰余,染色体末端会被连接在一起产生具多着丝粒的复合染色体或环状染色体,这样不能正确分离(见第4章染色体突变)。这两种潜在致命的过程被**端粒**(telomerse)所防止,它是一特异的结构以复制独立的方式加在染色体末端。端粒还作为联会起始者的作用。它们与核膜相互作用并且是染色体中最早配对的区域。在酵母与锥虫中,亚端粒DNA在基因表达调控中起重要的作用,它们含有信息的沉默拷贝,通过非交互重组转移到满足表达的位点(参见交配型转换,抗原转换)。

端粒DNA包括短的串联重复序列。在很多真核生物中都鉴别到这种序列,并且往往是富含GC对,一条链上簇集着鸟嘌呤,另一条链上簇集着胞嘧啶(表5.2)。它们可以通过鸟嘌呤之间的非常规相互作用形成异常的四倍体结构,这可能保护端粒不发生末端连接反应(见第16章核酸结构)。在不稳定的线状质粒的末端添加端粒重复序列导致了如同着丝粒存在同样的稳定性。

表5.2 不同真核生物中的端粒重复序列

生物体	端粒重复
四膜虫	TTGGGG
锥虫	TAGGG
酿酒酵母	$(TG)_{1-3}$TGGG
拟南芥	TTTAGGG
哺乳类	TTAGGG

端粒由一特异的核糖核蛋白复合物被称为**端粒酶**(telomerase)加到DNA末端上,它由多条多肽链与一单个RNA分子组成,RNA分子包含两拷贝的在端粒DNA中出现的胞嘧啶富集的序列。端粒酶的蛋白组分具有反转录酶活性,但其活性只限于应用端粒酶特异的RNA作为模板。基于这个信息,现有模型认为端粒重复是通过一引物延伸/模板转换的策略添加到已存在的端粒3′末端,如图5.2所示。最远端的端粒重复被认为可形成能封闭端粒末端的结构,这样就可以防止不正常的末端连接。这可能还涉及DNA的环化和/或端粒结合蛋白的聚合。末端DNA的环化可以启动富含G的DNA链的合成。

端粒酶的成员与其他与端粒活性有关的蛋白质可以通过分析影响端粒功能的突变来辨别。酵母中发现有一种株特异的特征性与端粒长度有关的一个突变,具老化表型,通过这种突变鉴别出了*TLC1*基因(编码端粒酶RNA)和几个编码端粒酶多肽链和端粒结合蛋白如哺乳类的TRF1的*EST*位点(甚至较短的端粒)。老化的表型表明了端粒在指定细胞的寿命范围决定中可能起着重要的作用。这种推断被以下的观察结果进一步证明了:在特定的人类体细胞中端粒的长度随年龄增长而下降,而在生殖细胞中,端粒的长度一直保持着(另参见生长转换;见第18章癌基因与癌症)。然而在小鼠的寿命比人类短得多,而端粒的长度却比人类长,没有观察到与年龄有关的端粒缩短。有推断认为端粒长度缩短到一定限度会诱导一种DNA修复机制,并抑制细胞周期。如果这样,细胞扩增就要求端粒酶活力,在大多数人类主要的肿瘤中端粒酶活性的存在支持了这种假设;端粒酶可能是抗癌治疗的合适的作用物。然而,最近得到的端粒酶基因敲除的小鼠表现出与其野生型的对照相似的癌症发生率(见第26章复制)。

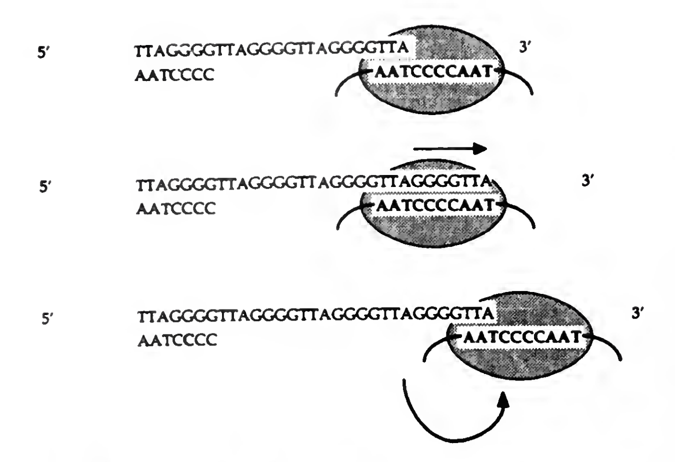
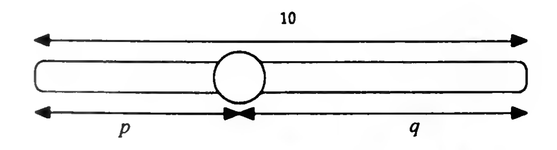


图 5.2 端粒酶活性的模型。(1)端粒酶 RNA(黑体)与染色体末端的 DNA 配对。(2)端粒酶 RNA 作为反转录的模板;DNA 引物延伸到模板的末端。(3)端粒酶转移到新的染色体末端。

在果蝇中,端粒的远端在结构上与其他真核生物相同并且可能以同样的方式——形成可封闭末端连接的帽子起作用。然而,端粒长度以一种新的方式得到保持,其中包括类 LINE1 逆转录因子特异地重复转位到染色体末端。这些元件定位的方式仍不清楚。

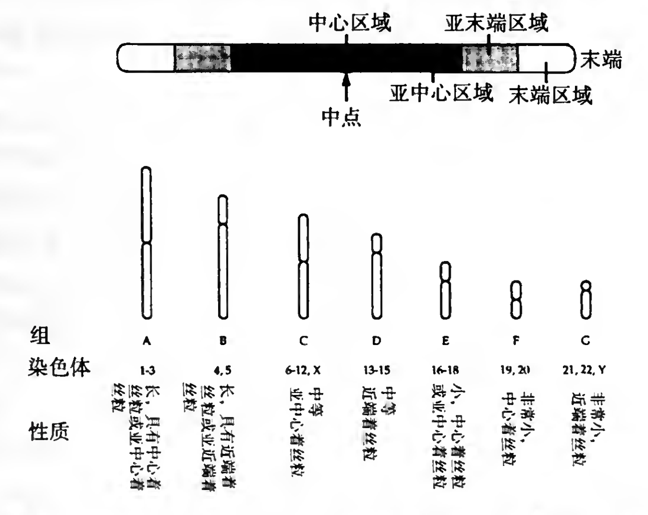
文框 5.1 染色体分类与命名

染色体参数 所有的已知的真核生物的单倍体都具有两条以上的染色体。所以必须定义一描绘性的命名系统可使每一染色体与核仁中的其他染色体区分开来。下面简要介绍这种命名法,它可特别指出着丝粒的位置与类型以及染色体臂的相对长度。大部分染色体可用三个参数来定义,d-值,r-值与着丝粒指数(i)。着丝粒指数限定了着丝粒出现的染色体区域而不考虑染色体长度。**d-值**(d-value)指染色体长臂与短臂之间长度的差别,即 $d = q - p$,q 和 p 分别指长臂与短臂的长度,并且染色体的总长被分成 10 个相对单位。**r-值**(r-value)指染色体两臂长度之比,即 $r = q/p$。**着丝粒指数**(centromeric index)指着丝粒到短臂末端的距离,用占染色体总长度的百分比来表示,即 $i = 100p/(p+q)$。根据这些参数与染色体的大小,人类染色体被分成几类,如下图所示。



图注:用于定义形态参数 d,r 与 i 的测量

术语	定义
单着丝粒的(monocentric)	具有单个确定着丝粒的染色体。大部分正常染色体是单着丝粒的。
全着丝粒的(holocentric)	具有弥散的着丝粒的染色体。因而整个染色体都发生纺锤体的附着(**外侧附着**,lateral attachment),如同在鳞翅目中发现的那样。
端着丝粒的(telocentric 或 telomictic)	具有末端着丝粒的单着丝粒染色体($p=0, q=10, d=10, r=\infty$以及 $i=0$),也被称为 **T-染色体**(T-chromosome)或**单臂染色体**(monobrachial chromosome)。这种染色体极少天然发生,大部分得到的单臂染色体都具有一较短但确定的 p 臂,因而这个术语被广泛误用了。它可用来描绘从着丝粒处断裂的畸变染色体。
非端着丝粒(atelocentric 或 atelomictic)	具有非末端着丝粒的单着丝粒染色体(**双臂染色体**,dibrachial chromosome)。进一步的分类见下。
中着丝粒的(metacentric)	具有中央或靠近中央的着丝粒的单着丝粒染色体。如果着丝粒正好位于中点,它被归类为 **M-染色体**(M-chromosome, $d=0, r=1.0$ 和 $i=50$),而如果着丝粒不是正中的而是位于**中间区域**(median region)中,就被归类于 m-染色体(m-chromosome, $d=>0\sim2.5, r=>1.0\sim1.7$ 和 $i=37.5\sim<50$)。天然的 M 染色体是极少的。(参见**等臂染色体**)
近中着丝粒的(submetacentric)	具有位于**亚中区域**(submedian region)的着丝粒的单着丝粒染色体($d=2.5\sim5.0, r=1.7\sim3.0$ 和 $i=25\sim37.5$)。
亚近端着丝粒的(subacrocentric)	具有位于**亚末端区域**(subterminal region)的着丝粒的单着丝粒染色体($d=5.0\sim7.5, r=3.0\sim7.0$ 和 $i=12.5\sim25$)。
近端着丝粒的(acrocentric)	具有位于**末端区域**(terminal region)的着丝粒的单着丝粒染色体($d=7.5\sim<10, r=7.0\sim\infty$和 $i=>0\sim12.5$),也被称为 t-染色体。很多被称为端着丝粒的染色体事实上是近端着丝粒染色体。



图注:用上面讨论过的命名法进行人类染色体分类。人类常染色体根据大小编号,1 号(最大),-22 号(最小),虽然事实上 22 号染色体较之 21 号大。

(徐　沁 译)

第 6 章
发育,分子机制

基本概念和定义

- **发育**(development)是复杂、多细胞生物产生的过程。它经常是从一个细胞开始,即一个受精卵细胞,但在特定物种中,它也可以从无性生殖开始,如通过单性生殖或从成体出芽开始。
- 发育是**渐成的**(epigenesis),是部分的变化过程,而不是**先成的**(preformation),即不是预先构建的小形有机体的增大。这样在早期胚胎形成简单和粗糙的模式,随胚胎生长不断完善。
- 发育可以被描述成三个重叠和相互依赖的过程。**分化**(differentiation)是细胞对应其不同功能特殊化的过程,反映了不同蛋白系列的形成。**区域专一化**是细胞获得**位置信息**的过程,此信息也就是控制它们行为和使它们在正确位置形成特定结构的指令,是**图式形成**(pattern formation)的基础。**形态发生**(morphogenesis)形成形状,反映了不同方面细胞的行为:生长和分裂,相对其他细胞的运动,形状和大小的改变、与其他细胞和细胞外基质的相互作用及细胞死亡。
- 除了少数例外,发育中生物的细胞保留相同的基因组,也就是发育主要是通过基因表达的调节来控制的。
- 发育过程是由内源和外源因子控制的。在大多数生物中,分化和图式发生的第一个相位是通过卵细胞中已经存在的(内在)不对称性。进一步的变化和图式形成是通过外源加工完成的——细胞-细胞联系和细胞间及细胞与环境间的相互作用。
- 发育从受精开始,但很难说明它是何时停止的。很多细胞变得是终末分化的,此被认为是它们的最终形式,而其他细胞——干细胞——并不是如此;很多发育过程是反复的。有些人把衰老看成是发育的过程;当生物死亡时发育才最终结束。

6.1 分化

基因组等同性 在分化过程中,细胞变得特殊化,以行使它们特定的功能,这反映在合成的一系列特殊蛋白的性质。除了少数例外(文框 6.1),发育中生物的所有细胞的基因组都是相同的,表明分化的基础是选择性基因表达。基因表达可以在很多水平被调节(见第 9 章基因表达与调节),一些发育基因调节的例子在表 6.1 中列出。

基因组等同性的证据是同时从分子和功能的研究中得到的。基因组 DNA 的分子分

表 6.1 在基因表达的不同水平上分化调节

基因表达水平	发育系统
转录	肌细胞的分化是由转录因子如 MyoD1 的合成来控制的,它们激活肌肉特异性基因。
RNA 加工	果蝇的性别分化是由性别致死、转化子和两性基因的基本转录本的选择性剪接来控制。 在蛾(*Manduca sexta*)的卵母细胞中很多 mRNA 的过早翻译因延迟的加帽而被阻止。
RNA 周转	鸡的肢芽发育部分是由 *FGF-2* 的表达控制的。与 *FGF-2* mRNA 互补的反义 RNA 在鸡的肢芽表达,目标定到 *FGF-2* 的 RNA 上,并使其降解。
蛋白合成	在果蝇早期发育前和后结构的建立部分是由翻译抑制控制的。Bicoid 蛋白抑制前部 *caudal* mRNA 的翻译,而 Nanos 蛋白在后部抑制驼背基因(*hunchback* mRNA)的翻译。 非洲爪蟾成纤维生长因子受体的合成在卵母细胞中通过蛋白与 *Xfgfr1* mRNA 的结合被抑制;此抑制在成熟的卵细胞中消失。
蛋白功能	果蝇胚胎中背-腹极性的建立依赖于 Dorsal 蛋白转移到细胞核,这是通过对抑制蛋白 Cactus 的磷酸化(从此失活)产生的。

析(如 Southern 杂交或 PCR)显示在所有细胞中无论其基因是否表达,基因组结构是完全相同的。功能实验显示分化的细胞在特定的环境中,能够**去分化**(dedifferentiate),而沿着另一条不同的发育途径进行,这样的现象称为**转分化**(transdifferentiation)或**组织转化**(metaplasia)。这表明在特定细胞类型中不使用的基因是潜伏的,能够被重新被激活。全能性的测试是从单个分化的细胞重新产生整个有机体。这在植物中是完全可能的,几种物种的分化细胞可以在培养中去分化,生长成为没有结构的未分化细胞克隆,称为**愈伤组织**(callus)。愈伤组织可以在植物激素的作用下,重新形成完整的植物,这样的过程被应用于有重要商业价值的植物物种进行遗传操作中(见第 24 章重组 DNA 与分子克隆)。对分化的动物细胞不可能得到同样的结果,但可能通过将两栖动物细胞核移位到摘除细胞核的卵细胞中,产生存活的胚胎。尽管胚胎不能发育成成体,这些实验证实分化的动物细胞,不象植物细胞,不能完全反转它们的发育定型(即使被分离和置于异常环境中),细胞核保留了某些潜能。用从两栖动物胚胎细胞的细胞核进行同样的实验,可以发育到成体阶段,且来自特定动物预先培植胚胎的细胞核也能支持完整的发育。这些结果说明动物细胞核不断失去发育过程的潜能,有关原因在下面讨论。除了这些限制,最近通过转移完全分化的乳腺细胞核到去除细胞核的卵细胞中,克隆得到一个绵羊,第一次证实分化的动物细胞核保留了产生生物全部功能的所有信息。

分化的简单模型　　多细胞生物的发育是复杂的过程,因为它不仅涉及到细胞类型的变化,而且涉及它们精确的空间组成。单细胞生物也可对环境反应而分化,它们提供了涉及在多细胞发育过程同样原理的简单模型体系——由内部和外部因子控制的不同基因表达——但缺乏模式形成。最简单的发育模型可能是枯草杆菌(*Bacillus subtilis*)的胞子形成,枯草杆菌有两种类型细胞:胞子和母细胞,在营养缺乏时分化(文框 6.2)。酵母细胞被证实是转分化的简单形式,它们可以自发从一种交配型转换到另一种交配型,这个过程涉及程序化非互换的重组(参见交配类型转换)。粘菌(*Dictyostelium discoideum*)在单细胞存在和简单多细胞生物之间交替转换,可能是复杂生物发育最有益的模型。细胞聚集

显示了动物胚胎的很多属性，包括模式形成、区域特异化的简单机制和调节缺失部分的能力(文框 6.3)。

早期发育的模式和分化 在多细胞生物中，分化和模式形成的过程是不可分割的。发育的早期涉及不同谱系的分离(一个例子是脊椎动物中的三个胚层分化，即外胚层、中胚层和内胚层)和以主体轴形式建立第一个初始的位置轴。

正如上面所讨论的，发育过程由内源和外源两种信号控制的。在大多数动物和植物中，发育过程由卵细胞中内源生化不对称性所起始。当卵细胞分裂时特定的分子进入不同的细胞，因此由胚胎的不同位置所产生的细胞就不再是等同的了。这样的分子称为**胞质决定子**(cytoplasmic determinats)，它是母体基因组在卵细胞中所处特定位置的产物。因此是母体基因型决定了胚胎早期发育的表型，这个现象称为**母体效应**(maternal effect)(图 6.1)。

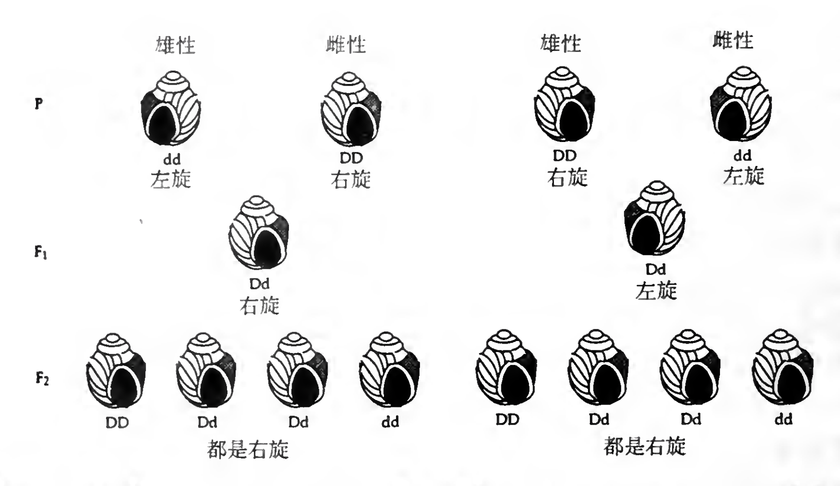


图 6.1 在椎实螺(*Limnaea*)中的母体效应。螺壳的旋转方向是由单基因座 *D* 决定的，显现等位基因 *D* 决定右旋，而隐性等位基因 *d* 决定左旋。*D* 等位基因编码的细胞质蛋白在卵母细胞中合成，它决定了第二次分裂中有丝分裂纺锤体的方向；这个因子在左旋卵母细胞中没有，说明左旋是缺陷途径。互交产生 F1 代椎实螺有确定的杂合基因型但螺壳旋转表型决定于母体在卵母细胞中对 D 蛋白的合成能力，也就是依赖于母体的基因型，而在每次杂交中不同。如果 F_1 代自交，F_2 代都是右旋，因为杂合的 F_1 椎实螺能在卵母细胞中合成 D 蛋白，说明这种右旋和它们本身的表型无关。注意：等位基因的分离是符合孟德尔原理的(参见孟德尔遗传)。

外源刺激可导致进一步的发育，在所有的脊椎动物胚胎中，如背腹轴是由环境中外源的物理刺激决定的。原肠胚从胚胎的背侧表面开始，这样建立了前后轴。在蛙中，胚胎的背腹轴是由受精行为决定的－精子进入的位置的对面区域经涉及皮质旋转的机制变成背侧。重力和旋转对在鸟类和(假定的)哺乳动物中对极性来说也很重要。在各种情况下，同源分子在背侧组织中心表达(文框 6.4)。

卵细胞中多少是由细胞质决定安排的，这有一个限制。在早期发育中，内源性指令用来作为多样性的唯一机制，而以后的特异化由外源指令控制，即这种指令是从细胞间相互作用中产生。在这种过程中，通过细胞质决定和物理刺激的表达分离建立初始的模式，可以经过细胞间相互作用对每一层的细节进行而不断精细化和补充。

镶嵌式和调节式发育 在早期发育中胞质决定和细胞之间相互作用的相对优势在动物界中是不同的。通常，胞质决定出现在大多数无脊椎动物胚胎，而细胞之间相互作用是在

脊椎动物发育中。

这两类信息来源决定了发育的两种极端机制,称为镶嵌式和调节式(图 6.2)。卵细胞或胚胎的**原基分布图**(fate map)显示了每个区域的**命运**(fate)(正常发育的结果)。在**镶嵌式发育**(mosaic development)中,单个细胞的命运完全由它内在性质决定(即它的胞质决定),而在**调节发育**(regulative development)中,每个细胞的命运由它与其他细胞的相互作用决定。在镶嵌发育中,每个细胞进行**自主特异化**,即在原则上,如果将它从胚胎中取出,它会根据内在指令发育。一个分离的细胞将会分化成胚胎的正确部分,而剩余的胚胎会缺失这一部分(参见*群体效应*)。胚胎的原基分布图可以从单个细胞在分离状态下的分化的命运来重建。在调节发育中,每个细胞进行**条件特异化**,即如果将它从胚胎中取出,它不能完成其最终命运,因为它将缺乏与其他细胞正确的相互作用。剩余的胚胎可以调节自身,替代缺失的部分。这是因为它们还会发生正确的相互作用,所以剩余细胞的结局还未被决定。驱动这样相互作用的胚胎区域称为**形成体**(organizer);它有一种特别的属性,那就是当转移到另一个胚胎区域中会形成新的结构,这是由于它是基本诱导信号的来源(图 6.2)。

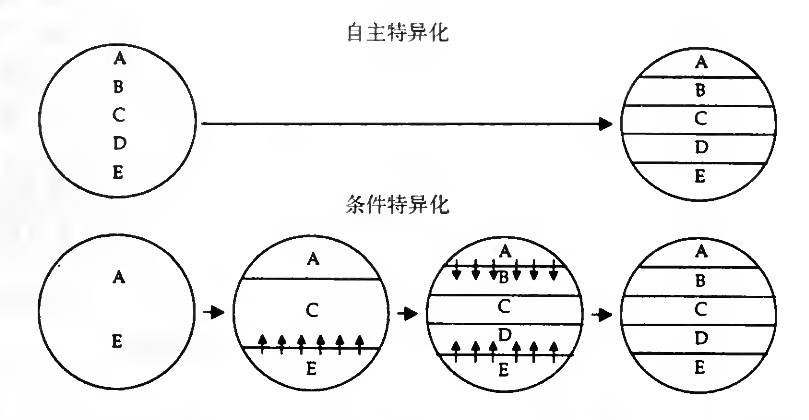


图 6.2 自主和条件特异化。在自主特化(上图)中,胚胎的五个区域(A~E)已经通过卵细胞中胞质决定特异化,并通过剪切而分离。在条件特异化(下图)中,胚胎的五个区域是通过细胞－细胞相互作用特化。两种胚胎都有同样的原基分布图,但对分离实验有不同的反应。如果对应于组织 B 的区域从早期胚胎中移去,上面的胚胎会显示镶嵌式发育(分离的细胞会分化成组织 B,而剩余的胚胎会缺失 B),而下面胚胎会显示调节发育(分离的细胞会分化成组织 A,因为它们还没有特异化为 B,剩余的胚胎会对缺失部分调节,也就是 E 会作为形成体诱导 A 形成 C,A 会诱导 C 产生 B)。

镶嵌发育的最极端例子是发生在有被膜的胚胎,它的主要胚胎细胞类型已经被卵细胞胞质决定特异化。在有些物种中,如 *Halocynthia roretzi*,决定的分布非常明显,因为它们是不同颜色的!即使这样,正确的发育需要有些诱导的相互作用,如神经系统通过卵裂球间的相互作用先于原肠胚形成之前而特化。哺乳动物提供了一个有调节发育的极端的例子。在卵细胞中没有内在的不对称性,在最初的几次分裂中,胚泡中的任何细胞都可以被分离,并发育成完整的胚胎。最早的分化是将滋养外胚层从内细胞团中分离出来(简而言之,它们将发育成胚外膜和胚胎)。这个决定看来是完全随机的,每个细胞的命运取决于它的位置。是在表面还是在胚泡中。大多数生物某种程度上属于这些极端例子之间,而出现在发育的早期发育阶段有自主和条件特化的例子。非洲爪蟾被用作例证(文框

6.4)。很多昆虫，如果蝇，在发育早期有独特的方式，它的特化发生是和合胞体有关的，细胞的命运在细胞形成之前被各个细胞核中可扩散调节因子的相互作用所决定。细胞形成后，进一步的特化通过细胞-细胞相互作用而发生。果蝇的发育在这方面被研究得特别清楚，在文框 6.5 和 6.6 中讨论。

发育中的细胞-细胞相互作用　细胞通过改变基因表达方式和细胞中已经存在的蛋白分子的活性对来自其他细胞和环境的信号作出反应(见信号传导)。这些反应都是分化的潜在途径，一个细胞(或者更正确地说是它产生的物质)影响了另一个细胞的发育命运，这样的过程称为**诱导**(induction)(见文框 6.4)。

诱导信号有多种形式。首先，可以区分涉及不同细胞类型和同等细胞的相互作用。在前一种形式中，细胞间的信号可以通过局部分泌的物质经过短距离(旁分泌)激活或者远处细胞释放的信号通过血管系统转运到他们的目标(内分泌)来介导。细胞-细胞的联系需要信号传导〔有时称为近分泌(juxtacrine)信号〕，或者细胞可能对分泌到细胞外基质中的分子作出反应。有两类诱导相互作用。在**指导诱导**(instructive induction)中，在诱导因子存在和缺乏的情况下，反应细胞采用不同的命运，即诱导细胞指导效应细胞遵循特定的发育途径。在**许可性诱导**(permissive induction)中，效应细胞已经注定某一命运，但需要特定信号使它们得以继续进行，即别无选择。对诱导的几种反应也可加以区分。经典的反应是相同的，也就是诱导信号引起效应细胞采用单一的、特有的命运。这经常发生在两层组织在一起时(附加诱导)或者被认为是在单细胞水平上诱导被考虑时。**形态发生素**(morphogen)是在给定细胞群(形态发生区域)中建立浓度梯度的一种可传播信号。**自体诱导**(homeogenetic induction)是可传布的反应，反应组织被再诱导分化成诱导组织，然后诱导它附近的细胞进行同样的反应。这些不同类型的诱导在图 6.3 中作了总结。

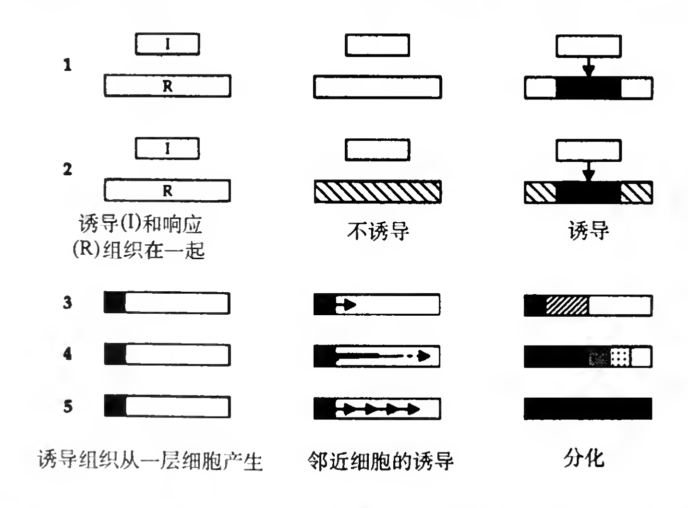


图 6.3　诱导的机制。上面的图表显示了许可性诱导和指导诱导的差别。在许可性诱导(1)中，反应组织的命运是固定的，但需要诱导信号来完成。没有其他选择，在没有诱导时，组织不能分化。在指导诱导中(2)，在诱导信号存在时效应组织以一种方式分化而在诱导信号缺乏的情况下以另一种方式分化。下面的图表显示了不同形式诱导信号的结果。在附加诱导中(3)，诱导细胞引起效应细胞产生单个反应。相反，形态发生素(4)是可传播的信号，在细胞范围中形成梯度，引起分化细胞根据它们的位置采取不同的命运。在自发诱导中(5)，诱导细胞效应细胞采取与诱导细胞同样的命运。效应细胞变成一个诱导因子，能够诱导邻近的细胞，信号可以在细胞群中传播直到它们都采取同样的命运。

上面所讨论的诱导机制都涉及不同细胞间的相互作用。分化也可以在相同细胞群中通过称为**边缘抑制**(lateral inhibition)的过程产生。在这种模型中(图 6.4),相同的、未分化的细胞分泌一种信号,抑制周围细胞的分化并抑制进一步的信号。在所有细胞信号以低剂量传递给它们邻近的细胞在此区域建立了平衡。偶尔产生的不平衡可以打断这样的信号过程,并被反馈所放大。中心细胞所释放信号的随机爆发将会抑制周围细胞的信号,而使中心细胞分化。然而,周围细胞不仅不再抑制分化,而且也不再抑制从中心细胞来的信号。这样中心细胞通过持续高水平的信号建立了优势,同时周围细胞既不能传递信号也不能分化。这样的机制被认为是果蝇外胚层和脊椎动物脑中神经元前体细胞分化以及线虫(*C. elegans*)子宫中锚细胞分化的基础(见文框 6.6)。

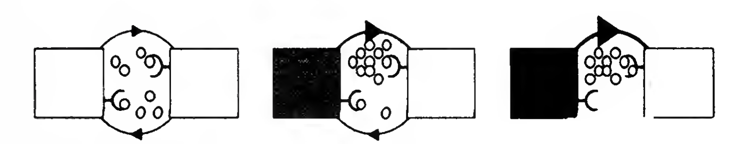


图 6.4 边缘抑制。在相同细胞群中分化通过信号分子来抑制,这些分子下调节周围细胞信号的产生,也可上调节受体的合成。起始的平衡通过相互的抑制维持,但在一个随机事件中,一个特定的细胞瞬间合成了比其周围细胞更多的信号,抑制周围细胞产生信号的能力,增加它们对信号的接受能力。在周围细胞最终停止信号产生,只接受信号的地方形成正反馈环路。在图中例举的果蝇神经发生中的 Notch/Delta 信号途径。

与边缘抑制相反的是**群落效应**(community effect),在这种效应中和分离状态细胞行为和在相同细胞群体中是不同的。这种效应是通过自分泌(细胞对其本身分泌信号有反应的能力)信号产生的。在这种情况中,相同的细胞分泌它们本身和周围细胞分化所需的信号。一群细胞协同的效果可以产生足以改变它们命运的信号,但单个细胞不能达到相同的反应。

在发育的过程中有很多诱导的例子,其中一些涉及早期主要的过程(如中胚层诱导和神经诱导,在文框 6.4 中讨论),有些发生在晚期特定器官的形成中。后者研究得较为清楚的例子包括线虫中阴门的特异化(文框 6.6)和果蝇眼发育中 R7 光受体细胞的特异化(文框 6.7)。

分化状态的定型和维持　　通过不断的细胞－细胞相互作用胚胎的变化使开始少量的细胞类型能够精细地分化成成熟动物或植物中很大细胞类型。这以一个**级系**(hierarchy)发生:当细胞类型产生越多,诱导相互作用类型也可能越多。在本章开始所讨论的动物细胞的一个显著特点,是它们倾向于定型在特定的发育途径上。一旦早期脊椎动物胚胎形成外胚层、中胚层和内胚层,这些细胞系其细胞的命运就被限制,只能分化成它们自己细胞系的细胞类型,而不能是其他细胞系。例如,一旦一个细胞作为外胚层分化,它的命运就限制在表皮或神经组织,它不再能使发育途径成为中胚层的延伸,如肌肉。动物发育的过程这样沿一系列不可逆的决定进行,细胞不断限制在越来越狭窄的命运中,直到它们最终分化。在这点上,细胞停止生长(如神经元,肌肉)或只产生它们同样类型的细胞(如角质细胞)。这个原则的主要例外是干细胞和胚细胞(见表 6.2)。

表 6.2 与细胞特定发育途径定型相关的名称

名称	定义
定型(commitment)	导致细胞沿特定发育途径的一种性质
去分化(dedifferentiation)	反转发育途径的能力,从分化的状态变成更易变化的状态
决定(determination)	细胞对特定命运的不可逆定型,是驱动细胞进入同样命运与其环境状态无关的条件
分化(differentiation)	通过合成或激活一系列新的蛋白而采取新的表型。**终末分化**(terminal differentiation)是在任何特定细胞系中的最后状态,细胞变得静止或只产生同样类型的后代(参见干细胞、胚细胞)
命运(fate)	特定细胞在将来会成为分化细胞类型。**原基分布图**(fate map)是显示卵细胞或胚胎所有细胞的命运的图谱
生殖细胞(germ cell)	形成配子的细胞。生殖细胞在动物发育早期经常从体细胞中分离,是在成熟动物中唯一保持全能性的细胞。相反,植物生殖细胞在成熟植物中从体细胞而来。生物学上的基本差异可以解释分离的动物和植物细胞有不同的潜能
潜能(potency)	一个细胞所有可能命运的总和(在任何环境中、不只是在正常发育中)。**全能性**(totipotent)——能够采取任何命运,也就是原则上可以形成完整的生物;**多能性**(pluripotent)——能够采取几种,但不是所有命运
特化(specification)	在中性环境的分离状态下细胞对特定命运的定型。确定通常的发育程序,但可以被外在影响所改变
干细胞(stem cell)	在每次分裂中产生两种类型派生的细胞:一个是其本身的拷贝,一个会分化(参见胚胎干细胞)
转决定(transdefermination)	定型的改变偶尔可以在果蝇成虫盘细胞中看到的,通过移植到果蝇成虫的腹部可维持生长状态,而不是进行分化。这样的变化不是随机的,但以预定程序发生,可能反映了调节蛋白活性的微小改变
转分化(transdifferentiation)	去分化和改变发育途径的能力。经常在切除后再生组织中发生

定型发生在几个阶段。一个细胞可以通过内在性质,采用特定的命运而被**特化**,在分离状态下以这种途径分化。然而,当暴露于不同环境的区域,它的命运可以通过与其他细胞的相互作用而改变。当一个细胞定型是不可逆转时,此细胞已被**决定**了,即它将沿着同样的发育途径进行,而不受环境影响。决定的时间选择可以通过移植实验来确定。爪蟾胚胎的外胚层特化成表皮,但将它移植到背中胚层也可以诱导形成神经板。在决定前,预定的表皮或预定的神经板在分离状态下都分化成表皮,但一旦预定的神经板被决定,它就变成神经板而不受环境影响——无论在原位、分离或移植到异位位置。特化和决定经常在分化之前,因为它们涉及激活控制分化的下游基因的调节分子合成。两个定型状态之间的差异反映了调节分子它们本身是如何被调节的。决定涉及调节环路的永久维持,赋予细胞对定型状态的记忆。维持可以通过细胞质反馈回路而发生或对染色质结构的识别而发生,这些机制可以帮助解释早先的核移植资料。在肌肉细胞系中,决定被反馈所控制。肌肉分化是被称为 MyoD 家族的 bHLH 调节蛋白所控制。在体节中,一旦这些调节因子在体细胞中被可诱导的相互作用激活,它们可以激活肌肉专一性基因的转录(如骨骼肌动蛋白和肌球蛋白),起始肌肉的分化。而且 MyoD 也激活其本身基因和其家族中的其

他成员，并通过正反馈维持它们的表达。一旦细胞开始表达 MyoD 调节因子，它变成不可逆的定型肌肉细胞系。然而，在原则上，细胞质环路可以通过将细胞核移植到其他细胞质中而被阻断。相反在其他细胞中定型状态的维持发生在染色体本身的水平上。在果蝇中，同源异型基因的表达是通过染色体结构的重建维持的，通过多梳（Polycomb）和三胸（trithorax）基因产物调节（文框 6.8）；在雌性哺乳动物中，染色质结构和 DNA 甲基化使一条 X-染色体维持在失活状态（见 DNA 甲基化和基因组外调节）。这种形式定型状态的维持不能通过细胞核移植到不同的细胞质环境中重建，而在动物中流行的基因组外定型可以解释体细胞核不能起始去核卵细胞的发育。植物中成功的体细胞克隆表明定型的状态是被外在加工所维持，也就是细胞 - 细胞之间的信号传导作用，而不是内在的调节系统，所以细胞转移到不同的环境中时仍能足以重建发育的程序。类似的情况可以在动物细胞中看到，细胞对受伤反应而去分化，并促进再生。在这种情况中，分化在没有决定时发生，使分化状态被所处环境所控制。

6.2 模式形成和位置信息

模式形成　　**模式形成**（pattern formation）是细胞采用一个精确的空间组织过程，保证物种所有成员在形态上类似。模式形成是由区域专一化和形态形成所组成。**区域专一化**（regional specification）是细胞获得**位置值**（positional values）的过程，也就是分子信息导致细胞在整个生物中以正确方式转运到它们的位置中，通过打断位置信息和产生相应的行为。

区域专一化　　区域专一化可分为两种形式，一种是控制分化并导致在特定位置产生特定的细胞类型，另一种是在分化状态相同细胞中控制形态形成，也就是导致它们行为的差异，产生区域专一化的结构。

这两种过程的例子在脊椎动物肢的发育中却可以发现。肢由很多不同细胞类型组成，有的形成骨骼、有的形成肌肉、有的形成皮肤等，由这些细胞组成一个骨骼在内，肌肉包在外，皮肤在表面的近似同心圆的模式。一旦这个过程开始，在不同位置上的相同细胞，如在前肢和后肢之间，沿肢三个主要的轴的细胞间产生结构上的分化。这样的差异反映了每个位置上细胞的不同行为，如不同的生长和形状的改变。

位置信息　　细胞获得它们的位置值的方式是通过基因表达模式所赋予的**位置信息**（position information）。位置信息的来源首先是在果蝇中通过分析功能破坏的身体模式突变（表明特定细胞获得了错误的位置信息）被证实的。两个系列的基因控制沿前后轴的位置值，**体节极性基因**（segment polarity genes）的作用是在每个体节中指定特定细胞类型的位置，而**同源异型选择者基因**（homeotic selector genes）的作用是指令这些细胞产生正确区域结构的方式来表现。特定细胞中综合的基因表达产生一个“地址”，并将其分配到胚胎的特定区域。体节极性基因的突变引起体节特定区域的缺失，它们被余下细胞的镜像所替代（也就是细胞类型不是相应的特异），而同源异型选择者基因的突变是奇异的，有时是奇怪的表型，身体的一部分被另一个类似物所替代（也就是正确的细胞类型是存在的，但它们的表现反常）。在果蝇中这些基因的调节和功能在文框 6.8 中讨论。

引人注目的是体节极性基因和同源异型基因在动物进化中都是保守的,并在所有动物的模式形成中起重要作用,这包括脊椎动物。脊椎动物同源异型基因的作用和它们与果蝇同源异型的类似性在文框 6.9 中讨论。肢的发育(文框 6.10)提供了两个系列基因如何在脊椎动物发育中发挥功能,不仅这些基因本身的功能是保守的,而且它们调节的相互关系也是保守的。

动物的同源异型基因在植物界中没有发现:完全不同的身体设计和发育机制表明完全不同的调节策略。然而,根据动物所拥有的 MADS 盒(在哺乳动物血清反应因子中也有的基序)的转录因子家族在拟南芥属(*Arabidopsis thaliana*)中发现,并且它们的功能缺失引起在花的重复结构的(花瓣、萼片、雄蕊和心皮中)同源异型转化。似乎相似的原理控制了这些因子和动物同源异型基因的表达和活性,因此 MADS 系统可能是植物中位置信息的基础。但是,有关植物中这些调节因子和这些分子的效应物以及细胞 - 细胞信号还有待于研究。

发育中形态发生的作用 形态发生是形式和结构的产生,可以被看作是发育的"第三"个成分,作用于已分化并赋予位置值的细胞,将它们所拥有的指令翻译成特定的运动和在空间上产生精确排列的行为。然而,除了形态形成确实作为发育程序的效应作用以外,它还在控制上起很强的活化作用。从发育的非常早期的阶段,细胞行为的方式——它们如何生长和分裂,如何相互黏附,如何改变形状——在胚胎中控制细胞的相互作用然后建立分化和区域专一性的框架。一个基本的例子是原肠胚的形成,在原肠胚中动物胚胎的结构在发育早期基本上被重新组成:这涉及到很多形态形成的过程,包括细胞形状的改变、细胞增殖和通过细胞黏附分子的表达方式的改变所引起的细胞亲和性的差异。在文框 6.4 中显示爪蟾胚胎原肠胚形成的结果 - 背部中胚层被赋予位置信息,因其可诱导所以置于内胚层之上,变成神经板。因此形态形成可作为分化和区域专一化后驱动力量。

形态发生的机制 形态发生的基础是细胞行为。这可以被认为是独立的过程或根据细胞与环境的相互作用的过程。细胞内在的驱动形态形成过程包括细胞生长和增殖的速率,细胞分裂的性质,细胞外形的改变和细胞死亡。在表 6.3 中显示了这些在发育中发生的过程。对有些物种这些过程是精密调节的。在线虫纲的蠕虫(*Caenorhabditis elegans*)每个细胞的分裂和死亡都有恒定的发育程序。可预计的移动意味着所有细胞 - 细胞诱导性相互作用也是不变的,每个个体在精确的位置上含有同样精确数目的细胞。在 *C. elegans* 鉴别出了多个调节这种细胞行为的基因。一个例子是 *lin -4* 基因,它编码的一个转录因子在幼体的第一阶段所需要的。LIN - 4 蛋白水平上的修饰通过引起细胞行为以异时(heterochronic)方式——即符合早先的或后期的发育阶段,阻断几种细胞系。*lin -4* 过度表达导致很多细胞表现为发育早期阶段不成熟的和重复分裂的性质。功能缺失突变相反地引起细胞跳过某些分裂,产生更符合后期发育的形式。在更为广泛的生物体中,细胞生长和细胞分裂倾向于整个被生长因子和激素所调节,而不受内在的程序控制。这在细胞数目和统计资料上有很大的变化而不是程序化地一个给定细胞确定的分裂。这样的随机影响可能在发育噪音(参阅)中起主要作用,发育噪音是在恒定环境中维持同基因型群体变异的来源。

表 6.3 在发育中形态发生的过程

形态形成的机制	发育中的例子
细胞生长(大小增加)	神经元对神经生长因子的反应
增殖	在两栖动物原肠胚形成中的织原(organizer)细胞
细胞分裂	爪蟾卵不均等的分裂产生大的植物性分裂球和小的动物性分裂球
细胞形状改变	神经管关闭时神经板铰链区域的神经细胞顶柱变窄
细胞融合	肌管从成肌细胞中分化
细胞死亡	很多神经元
	趾间坏死区域中间充质细胞的死亡
细胞-细胞黏附消失	哺乳动物和鸟类:当细胞通过原条移动时细胞分层
细胞-细胞黏附获得	在肢发育中软骨充质细胞的浓缩
细胞-基质相互作用	神经嵴细胞的迁移
细胞-基质相互作用消失	从皮肤基层刺激分化成角质细胞的分层

驱动细胞形态发生的外在过程包括细胞-细胞黏附和细胞-基质相互作用。细胞可以改变它们表达的细胞黏附分子的内容,这样来改变与周围细胞的关系。两个细胞黏附分子家族已被识别:Ca^{2+}依赖的钙粘着蛋白和免疫球蛋白超家族的 Ca^{2+}不依赖的细胞黏附分子(如 N-CAM)。通过表达不同的 CAMs 和钙粘着蛋白,细胞可以选择是否相互黏着并可以相互移动,使它们形成片或块状结构或者分散。这样的相互作用对原肠胚形成中所产生的复杂的形态发生移动很重要,随细胞建立和打断联系胚胎中细胞黏附分子的表达也可以看到变化。细胞也通过细胞表面称为整合素的受体与细胞外基质中的层粘连蛋白和纤粘连蛋白作用。这种的细胞-基质相互作用对迁移细胞如神经嵴细胞和原生殖细胞的以及对角质细胞都很重要。

6.3 发育的环境

发育中的基因和环境　在前面的部分中,发育作为基因组中程序化的一种遗传过程,在细胞质中正确的设置下产生整个生物。在脊椎动物发育早期物理指示的应用明明需要考虑环境的整个作用。在群体中观察到环境影响在发育过程中引起的表型变化,以及发育噪音(参阅)背后的随机影响。像很多发育过程一样,性别决定同时涉及遗传和环境的因素,这个过程的特点是这些因素可以分开,且在不同的器官中所起作用的相对重要性是不同的。在不同的物种中存在初级性征(产生配子的类型)的发育和第二性征(个体的表现)是通过一些共同成分的不同途径调节的。在很多动物中一个重要的性别决定特征是性染色体(参阅)的存在。这些染色体在性别中的分布是不对称的,剂量补偿机制对矫正一种性别相对另一性别有两倍剂量的特定基因产物是必须的。在果蝇和哺乳动物中,雌性有两条 X 染色体,而雄性只有一条。在果蝇中,X-剂量补偿是通过在雌性中将 X-连锁基因座转录率降低到 50%(表 6.4)。在哺乳动物中,同样的结果可以通过使 X-染色体中的一条失活来完成(见 DNA 甲基化和基因组外调节)。

表 6.4 在性别决定机制中一些遗传优势和环境优势的例子

机制	描述
遗传优势机制	
哺乳动物——染色体决定	在哺乳动物中,初级的性征(性腺生理学)是由 Y 染色体上 *SRY* 基因的存在决定的。有 Y 染色体的个体经常是雄性,不管 X -染色体的数目有多少,而 XO 个体是雌性。*SRY* 编码一个含 HMG 盒的转录因子,诱导其他调节因子的表达(包括 SOX9 和 SF1),在性腺中可以抑制 Wnt - 4a 信号。SF1 在两种性别性腺发育早期都需要,但不存在 SRY 的雌性中其水平逐渐下降。在雄性中 SF1 表达得到维持,它在支持细胞中诱导副中肾抑制物质和间质细胞中睾酮的表达。这两种激素刺激中肾管(将来成为输精管)存留和副中肾管(将来的雌性生殖管)的退化。SRY 去除 Wnt - 4a 的效应可能是抵消 X -连锁的 *DAX* 基因的效应。当 *DAX* 存在一个拷贝表达不足以取消 SRY 的效应,但两份 *DAX* 基因座引起 XY 个体发育成雌性。 哺乳动物的第二性征决定于激素。在这种情况下,雌性是默认状态。缺乏睾酮受体时,XY 个体(内部)雄性生殖腺发育但是雌性的第二性征也得到发育。同样原理,雌性小牛在子宫中可被雄性孪生小牛产生的睾酮诱发产生雄性特征,产生称为**双生间雌**(freemartin)的动物。
果蝇——平衡机制	在果蝇中,性别基本上是由常染色体和 X -染色体之间的平衡决定的,而 Y -染色体为雄性生育所需,它不是雄性性征的初级决定。因此 XY 个体是雄性,而 XO 个体是不育的雄性。X:A(常染色体)比例为 1:1 时是雌性,而比例为 1:2 的是雄性,中间比例有中间表型,称为**雌雄间性**(intersexes),更极端的比例显示了极端的"雄性"和"雌性",称为**超雄性**(metamales)和**超雌性**(metafemales)。 平衡机制反映了 X -染色体和常染色体编码的不同螺旋-环-螺旋转录因子相对水平。X -连锁基因 *sis-a*(*sisterless-a*)、*sis-b*(*sisterless-b*)和 *sis-c*(*sisterless-c*)编码的因子,它们激活 *Sxl*(*Sex-lethal*)基因。而常染色体 *dnp*(*deadpan*)基因编码一个抑制物,可能通过阻止上述螺旋-环-螺旋成为非活性的异源二聚体。在 *Sis* 因子占优势的雌性中 *Sxl* 是激活的,而在雄性中,由于相对高水平的 *dnp*,*sxl* 这个基因是抑制的。*Sxl* 蛋白行使两个功能。第一是起始一个不同的 RNA 剪接反应,产生雌性专一性的 *dsx*(*doublesex*)基因转录本,第二是抑制一系列调节 X -连锁基因雄性专一性转录速度的基因。果蝇性别决定中 RNA 剪接反应在图 27.4 中显示(见第 27 章 RNA 加工)。
环境优势机制	
海龟——温度依赖	大多数爬行动物性别决定机制类似哺乳动物的染色体,但海龟和乌龟的性别是依赖温度的。这反映了将睾酮转化为雌激素的酶合成或活性是温度敏感的。酶活性本身受温度影响,或者依赖于调节酶合成的上游因子。
蜗牛——基质依赖	履螺属的蜗牛(*Crepidula*)形成含有 5～10 个体的丘状隆起,这些个体开始都是雄性的。在老的蜗牛中,生殖系统被破坏,根据基质重新产生雄性或雌性。在幼年的雄性群体中,如果与死亡的蜗牛接触就产生雌性,与雌性蜗牛接触产生雄性。在缺少的状态时是雌性——分离状态经常是雌性。

文框 6.1 发育中基因组的不均衡

随机和程序化改变　大多数发育生物的细胞常常包含同样的遗传信息,差异可以以两种方式产生。第一是通过突变,关于改变的类型和发生的位置(见突变和选择)两者都是随机的。第二是程序化的改变,改变类型和位置在发育过程中是严格地受到调节的。这些过程都涉及 DNA 获得、丢失或重排。

基因扩增　**基因扩增**是特定基因或其他 DNA 序列的拷贝数增加。在体内已识别有两种程序化扩增的模式:整个基因组扩增和选择性扩增。**整个基因组扩增**(whole genome amplification)通过增加细胞中基因组的拷贝数而增加特定基因的拷贝数。在真核生物中,在 DNA 复制后细胞没能进行有丝分裂,基因组就加倍,例如在哺乳动物肝细胞中是四倍体,但在双翅目(如果蝇)的分泌腺中最明显,很多轮的复制产生巨大的多线染色体(参阅)包含多于 1000 染色单体。**选择性扩增**是调节性发育过程偶尔用来增加需求量高的基因产物的量。有三个经详细研究的扩增体系。很多纤毛虫(如四膜虫)和有些多细胞动物(如蛙)选择性扩增它们的 rRNA 基因。在这两种情况扩增都发生在蛋白合成需求高的非常大的细胞中(如蛙的成熟卵母细胞),这反映了没有其他方法增加 rRNA 的产物——它不被翻译。在两种情况中,扩增都涉及从基因组中剪切 rDNA。原生动物有单个 rRNA 基因,而蛙在生殖细胞中以及有多于 1000 个 rRNA 基因。在四膜虫中,剪切的 rDNA 是通过发夹引导复制的,产生一个线性的反向重复元件。在蛙中,剪切下来的 rDNA 在扩增前环化。基因组的完整性在扩增中得到维持,因此似乎是额外的 DNA 环在剪切前进行复制,是重组引起单个 rDNA 单元被剪切。编码蛋白的基因很少被扩增,因为产量可以在蛋白合成水平上来提高。果蝇的卵壳蛋白基因是一个例外——卵壳蛋白在短期中大量需要,这通过在卵泡细胞中选择性扩增来满足。与 rRNA 基因不同,卵壳蛋白基因不被剪切,但在基因组中通过复制的重叠环而被扩增,产生一系列的**协同扩增子**(concentric amplicons)(**洋葱皮模型**)。在其他昆虫中,在生殖细胞中发现了多拷贝的卵壳蛋白基因,为什么在果蝇中进化成为一种选择性扩增现在还不清楚。

染色质减少和其他 DNA 缺失情况　基因扩增的同时,DNA 也可以是选择性的缺失或者涉及整个基因组的拷贝。在减数分裂时,所有细胞基因组都减少:在中间插入的 DNA 复制以外有两轮连续的分裂。一种特殊的 DNA 丢失的方式是从哺乳动物网织红细胞中细胞核被排除,分化成为无核的红细胞[**零倍体**(nulliploidy)]。选择性 DNA 丢失可以在很多低等真核生物中看到,它与体细胞和生殖细胞的分化有关。一般,生殖细胞保留了整套 DNA,而在体细胞系中有选择性 DNA 丢失或染色质减少。这个过程涉及位置专一性的 DNA 剪切和分离。其意义还不清楚,但看来是体细胞发育的重要方面,并被细胞质成分的区域性分布所调控。

程序化重组　少数细胞类型的分化依赖于重组。程序化非交换重组控制着酵母交配类型转换和锥虫的抗原转换,而程序化位置特异性重组控制着脊椎动物 B 细胞和 T 细胞分化的很多方面(见第 25 章重组)。

文框 6.2 枯草杆菌(*Bacillus subtilis*)的孢子形成

孢子形成途径概述 饥饿的营养期枯草杆菌细胞诱导孢子形成——基因组的一个拷贝分离成为前芽孢,外面包以芽孢外被,同时另一个保留在母细胞中,最终被破坏。这是分化的一种简单形式——一个细胞产生两个表型独特、功能不同的细胞。分化的过程主要是通过合成不同的σ-因子(细菌 RNA 聚合酶的成分,见第 29 章转录)和其他转录调节因子在转录水平上进行初步调控。孢子形成的途径涉及每个细胞中转录调节因子及交叉的转录后调节。

孢子形成的起始 当饥饿诱导的蛋白激酶级联反应在转录调节因子 SpoOA 被磷酸化时终止,此时孢子形成便开始了。激活的 SpoOA 通过与营养期σ-因子σA 和σH 的相互作用,诱导很多控制进入孢子形成途径的基因转录。两个新的σ-因子,σF 和σE 作为起始反应的一部分被合成。在从中间到两极的分离中有一开关,在细胞两个末端的两极位置形成蛋白 FtzZ 的环(参见**细菌细胞周期**)。只在细胞的一个末端形成隔膜,对应隔膜的小间隔区变成前芽孢。目前还不知道细胞是如何选择变成孢子的末端,但形成单孢子是依赖σE 的。分离后发生染色体分离,所以染色体必须穿越过隔膜进入前芽孢,这是由 SpoIIIE 蛋白执行的。

σ-因子的级联反应 孢子形成的过程可以分成很多依赖于不同系列基因表达,形态显著不同的阶段基因的表达与转录调节蛋白级联反应结构相一致。孢子分化的进行涉及**σF** 在前芽孢特异性的激活。σF 作为一种带有 SpoIIAB 蛋白的非活性复合物开始是均匀分布的。SpoIIAB 是磷酸化 SpoIIAA 蛋白的激酶;它的活性被膜结合的磷酸酶 SpoIIE 所拮抗。在它的非磷酸化形式时,SpoIIAA 可以与σF/SpoIIAB 复合物相互作用,替代σ-因子;因此去磷酸化的 SpoIIAA 的浓度原则上决定了σF 的活性。SpoIIE 被认为存在于隔片膜两侧,但因为前芽孢体积小,它的相对浓度高,使平衡向有利于 SpoIIAA 去磷酸化方向进行,结果在前芽孢中特异性替代和激活σF。

下一个阶段是关系到与在母细胞中σE 的特异性激活。与σF 一样,σE 开始以非活性形式均匀分布,但它的非活性形式是因为作为蛋白前体合成的,必须被蛋白酶剪切。蛋白酶是由 *spoIIGA* 基因编码的跨膜受体,它的蛋白酶活性被配体 SpoIIR 激活,SpoIIR 在σF 的调节下特异性在前芽孢中合成。因此σE 在母细胞中的激活需要前芽孢中的σF。SpoIIR 是从前芽孢中分泌到中间区域,激活剪切σE 的蛋白酶。σE 在前芽孢中本身是没有活性的,因为显然在 SpoIIE 蛋白的控制下,它被完全降解。

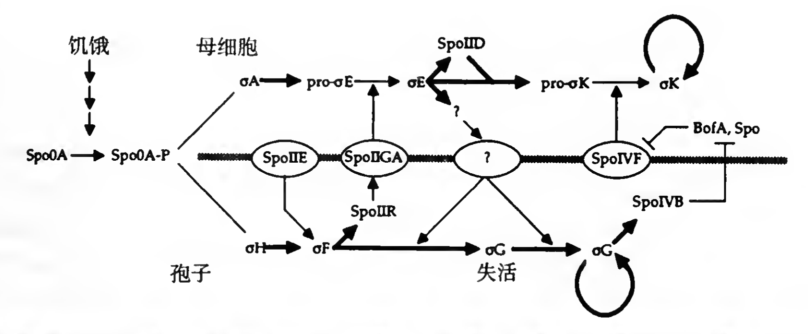
依赖σE 的基因在两个时期被转录:最初的基因包括 *spoIIID*,它编码一个第二时期σE 转录需要的转录因子。在第二时期的依赖σE 基因中有编码另一个σ因子σK 的 *sigK* 基因。σK 的表达调节非常复杂,因为这个基因被称为 *skin* 的隐藏的原噬菌体打断,*skin* 必须被切离。这个过程通过 *skin* 中编码的位置专一性重组酶 SpoIVCA 执行的,SpoIVCA 的表达依赖于σE 和 SpoIIID。一旦 *sigK* 基因座的两半重新结合,基因在σE 和 SpoIIID 的控制下转录,翻译产生一个类似于原σE 的非活性的前体蛋白。SpoIIID 可以作为σE 的激活因子和抑制因子,是σK 依赖基因。

在前芽孢中,σF 基因的进一步的调节在σF 本身激活后一小时被诱导。这包括了编码另一个σ-因子σG 的 *spoIIIG*,出现延迟反应了在母细胞中转录对σE 功能的依赖性(尽管所涉及的假定信号分子还没有被发现)。σG 激活也需要σE,尽管机制还不清楚。

在前芽孢中,σG 指导后期孢子形成基因的表达,其中一个是 *spoIVB*。它编码了涉及母细胞中σK 激活的信号蛋白。象原σE 一样,原σK 被膜结合蛋白酶激活(这种情况下是 SpoIVFB),它起先是非活性形式的,被来自前芽孢的信号所激活。然而,与正常情况下是无活性的 SpoIIGA 不同,SpoIVFB 始终是有活性的,但通过 *spoIVFA* 和 *bofA* 基因编码的抑制蛋白形成非活性复合物。因

此 SpoIVB 的作用是起抑制因子的作用，而不是直接激活蛋白酶。

σG 和 σK 分别控制前芽孢和母细胞的后期基因表达，都能正调节它们本身。另外，σK 依赖基因在两个时期被转录。基本的时期包括编码转录调节因子的 *gerE*，它激活和抑制一些其他 σK 依赖基因。



控制枯草杆菌孢子形成的调节网络。转录调节以细箭头表示，翻译后调节以粗箭头表示。膜蛋白以环表示。

文框 6.3 盘基网柄杆菌（*Dictyostelium discoideum*）的生命周期

生命周期概观　在营养状态下，盘基网柄杆菌以单倍体的单细胞存在，称为**变形粘菌**（myxamoebae），是无性繁殖的。引人注目的是，当单细胞消耗了营养供应时，它们聚集成串，迁移到中心点形成多细胞聚合体。聚集中的细胞分化，形成**前孢子细胞**和**前柄细胞**。前柄细胞在聚合体的顶部形成，而前孢子细胞在基部。聚合体在它的侧面伸出，移动时整个的聚合体就像一条**蛞蝓**[也称为**伪原生质团**（pseudoplasmodium）或**粘聚体**（grex）]。当蛞蝓从暗处移动到明亮的地方，它分化成子实体，由提高的柄及其上的一个孢子细胞组成。孢子细胞分散形成新的胶质变形体。

聚合体的分子基础　决定聚集是根据所提供的营养和细胞密度。胶质变形体可以通过监测称为**饥饿前因子**（prestarvation factor, **PSF**）的分泌蛋白来测定细胞密度。饥饿的细胞释放 cAMP 进入它们的环境中，如果 cAMP 和 PSF 的水平足够高，周围细胞开始聚集。起先，cAMP 由少量消尽营养的单个细胞释放。邻近细胞通过向信号的来源移动对此作出反应，而本身释放 cAMP，这样将信号传递到更外围的细胞。一旦一个细胞以这种方式反应，在它再次反应之前有一个决定的时期。因此信号作为不连续的脉冲从信号源散发出去。当它们迁移时，细胞开始合成新的细胞黏附分子，使它们可以形成柄，最终进一步聚集（可以包含多到 10^5 细胞）。

聚合体的细胞分化　聚合体的细胞分化成前柄细胞和前孢子细胞，这样前孢子细胞位于主体（变成蛞蝓的后部）而大多数前柄细胞位于顶端（变成蛞蝓的前部）。蛞蝓的后部也包含了分散的前柄细胞，称为**类前体细胞**（anterior-like cells, **ALCs**）。分化的过程被 cAMP 和相关的总称为**分化诱导因子**（differentiation indcing factor, **DIF**）的亲脂分子家族所调节。前孢子细胞形成需要 cAMP，因为 cAMP 诱导前孢子基因并稳定它们的 mRNA。前柄细胞分化需要 DIF，它诱导前柄细胞基因表达。细胞分选以几种机制发生：前柄细胞被认为分选成蛞蝓的前部，因为它们持续向 cAMP 来源迁移，而前孢子细胞不是这样。不同的细胞外基质蛋白的表达控制不同亚类的前柄细胞的分布。大多数前柄细胞表达 *ecmA* 基因，分选成蛞蝓的前部－那些高表达这些基因的细胞称为 PstA 细胞，占领了前部终极；那些低表达的细胞称为 Pst0 细胞，位于后部区域。在 PstA 细胞

中,锥形细胞群也表达 *ecmB* 基因。这些 PstAB 细胞最终起始柄的形成。最后,只表达 *ecmB* 的一群细胞发现在孢子细胞分散分布。起初选择哪些细胞成为前孢子,哪些成为前柄细胞可能依赖于生长停止和分化开始时的细胞周期的阶段:在 G_1 期阻滞的细胞比那些在细胞周期后面阻滞的细胞合成细胞黏附分子要少,结果在与其他细胞黏附前进一步迁移。

终极 子实体的构成称为**终极**(culmination),涉及蛞蝓中不同细胞群的重新组织,最终分化成柄细胞和孢子细胞。蛞蝓迁移在光照或低湿度下停止,这两种情况都引起氨(通过迁移的蛞蝓大量产生)分散更快。伴随着 cAMP 的产生,氨的消耗,释放了很多涉及分化基因的抑制。这些去抑制分化基因通过蛋白激酶 A 的活性,蛋白激酶 A 被认为磷酸化(和灭活)一个或多个抑制蛋白。终极的表现涉及提高柄的形成,它通过 PstA 细胞向早期终极的迁移和 *ecmB* 基因诱导转换成 PstAB 细胞完成。PstAB 细胞向下推进通过前孢子细胞进入基部,升高了孢子,这样它们可以被分散。这样做以后,它们形成空泡,合成纤维素;它们死亡并形成刚硬的柄。PstB 和 PstO 细胞也进行迁移和分化:前者向下迁移形成子实体基部,后者与剩余的 PstA 细胞向外迁移,形成孢子体和上下杯。

发育突变体 很多涉及盘基网柄杆菌发育的基因通过突变已被鉴别。这样的突变体是不能聚集的,因为它们缺乏发送、接受或反应 cAMP 信号的能力。其他突变体阻断了后期阶段,因此它们不能迁移、分化或对 DIF 或氨反应。还有其他分化异常,如通过缺乏特定的发育途径,打断正常的前孢子和前柄细胞的比例,或直接进入最后的分化途径。有多到 400 个基因涉及这些发育过程。

文框 6.4 爪蟾发育的早期事件

轴形成概观 爪蟾和其他两栖动物都表现了在发育的早期自发和条件性特异化。卵细胞在分子上和生理术语上都是不对称的,因此在受精前可以辨别**动物-植物轴**(animal-vegetal axis)。胚胎将来的背腹轴由物理线索决定-精子进入的位置。受精引起胚层的旋转,精子进入点的相对位置的细胞质的混合,在称为 **Nieuwkoop 中心**的区域诱导植物性细胞质中的背部信号。在分裂时,植物性细胞诱导上面的动物性细胞形成中胚层。Nieuwkoop 中心的细胞诱导有组织者活性的背部中胚层,而腹部的植物性细胞诱导腹中胚层。称为**组织者**(organizer)是因为它起始原肠胚形成和特异化胚胎的前后轴,当转移到一个异位时,也能这样进行。组织者也有很多进一步特异化的性质。第一,它可以分化成背中胚层结构(脊索、体节、头间充质);第二,在原肠胚形成前,它可以传递信号到邻近的腹中胚层,诱导边缘和中间的中胚层结构;第三,在原肠胚形成中,它可以使上面的外胚层背向化,也就是诱导它形成神经板,这样以后将其本身的前后位置信息输入神经组织,确定前后神经轴。这些相互作用的分子机制在下面讨论,并在相伴的图中总结。

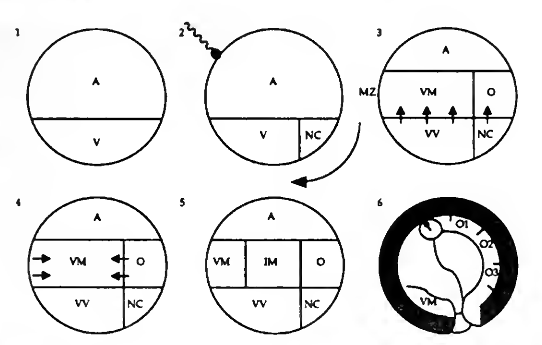
Nieuwkoop 中心的建立 Nieuwkoop 中心是通过胚层的旋转建立的,这看来涉及转录因子 β-连环蛋白在卵背部的积累。这个蛋白合成起先是贯穿整个卵细胞,人们认为受精位置诱导了 β-连环蛋白抑制因子糖原合成酶激酶-3 的活性,尽管其具体机制还不清楚。β-连环蛋白(β-catenin)的活性通过 TGF-β 信号传导途径增强,植物性位置的母体蛋白 Vg-1 被认为在体内起这样的作用。因此,激活的 β-连环蛋白通过受精的作用将限制在卵细胞将来的背侧,被植物性特异的细胞质决定所激活。这足以将 Nieuwkoop 中心定位于卵细胞的背腹四分之一位置。β-连环蛋白诱导对称为 Siamois 的转录因子的表达,Siamois 被认为是调节对 Nieuwkoop 中心功能重要基因的转录。

中胚层诱导 植物性细胞诱导上面边缘位置的动物细胞成为中胚层,纤维生长因子(FGF)和转化生长因子-β(TGF-β)家族的分子都与这个信号途径有关。这两类分子都出现在爪蟾胚胎中,通过过量表达显型失活受体分子使其受体去活性会严重破坏中胚层的发育。灭活类 TGF-β 因子活化

素受体完全破坏了中胚层的形成，而涉及FGF受体类似的实验结果是腹部中胚层衍生物丢失。这些结果说明类TGF-β因子在中胚层形成中是指令，但FGF是腹中胚层特异化所需的。Vg－1是TGF-β家族的成员，为Nieuwkoop中心的功能所需。在体外分离的动物组织中，增加Vg－1的剂量可以诱导更多背中胚层标记的表达，而在体内，还需要β-连环蛋白，这说明Vg－1可能与Siamois转录因子激活的未知产物作用一致。FGF信号可能因此调节胚胎腹侧对Vg－1的反应，特异化中胚层的细胞归宿。然而，泛素信号蛋白BMP－4似乎通过诱导腹中胚层特异的转录因子(见下文)在腹中胚层特异化中起直接作用。腹中胚层的诱导因此可能涉及Vg－1,FGF和BMP－4的组合。这些分子的相互作用对爪蟾是特异的。通过基因敲除(参阅)实验的结果，很多涉及的基因在小鼠中是保守的，但似乎起不同的作用。

组织者的分子控制　从Nieuwkoop中心来的信号诱导很多组织者特异的转录因子合成，这些激活的基因涉及到背中胚层的发育，将信号传递到邻近的腹中胚层、原肠胚形成的移动和随后的信号传递到上面的外胚层。很多转录因子特异性定位于组织者，包括XANF－1、Pintallavis、Lim1和Goosecoid，它们重叠的表达模式似乎可以将组织者再分成功能特异的区域。从组织者来的背部化信号诱导邻近的腹中胚层表达中胚层标记。这些信号似乎通过拮抗BMP－4(BMP－4在整个胚胎中都表达，可能是腹部化信号，诱导几个转录因子的活性——Xvent－1,Xom和Vox——它们不仅激活涉及腹中胚层分化的下游基因，而且抑制背部化的信号活性，如Goosecoid)的活性起作用。组织者至少合成三种分泌蛋白- Noggin,Chordin和Follistatin——看来它们的功能是抑制BMP－4信号。在Noggin和Chordin中是通过直接与BMP－4结合，而Follistatin是与BMP－4上游的BMP－7结合。BMP－4和它的抑制分子在胚胎中的浓度梯度的相反，这样特异化了邻近组织者的中胚层区域。

神经诱导和前后神经轴的特异化　神经诱导的分子基础还不清楚，但输入到神经板的位置信息分子已经被确定。Noggin,Chordin和Follistatin引起上面外胚层形成前神经结构(前脑)。人们认为后部的信号引起尾部化，这可能是FGF。第一进入胚孔的细胞合成一个称为Cerberus的分泌蛋白，它特异化头部最前面的结构。



1.事先已存在卵中的动－植物极的不对称性(A.V)。2.受精导致皮层旋转并通过β连环蛋白的区域激活，建立了Nieuwkoop中心(NC)。3.植物极细胞诱导了边缘区(MZ)的动物极细胞使其变成为中胚层。NC诱导了背部的中胚层(组织者,O)，有可能是通过Vg－1和转录因子Siamois所激活的产物二者协同作用。腹部的植物极细胞(VV)以涉及到FGF因子及BMP－4的一种途径诱导腹部的中胚层(VM)。4.来自组织者的侧翼信号(Chordin,Follistastin,Noggin)在腹部中胚层中与BMP－4相互作用。5.BMP－4分极抑制建立了一个中间胚层型的区域(IM)。6.组织者起始了原肠胚形成。中间中胚层被置换到胚的侧面，最前面的组织者细胞分泌了Cerberus蛋白，此蛋白诱导了头部前端的结构，如胶粘腺(C)。余下的组织者诱导神经板(N)；那些越过胚孔迁移细胞(O1)首先变成为前部的中胚层，并诱导前部的神经板(N1)。神经形成信号的特点还不知道，但它可以激活外胚层蛋白激酶C和腺苷环化酶。Chordin, Follistain和Noggin诱导前部神经的命运，次级信号也能是FGF常常可以结束神经膜，产生后神经结构(N2,N3)。

文框 6.5 果蝇中体轴的多核体特异化

概观 如上文所讨论的,早期昆虫发育是在合胞体环境中发生的,合胞体在原肠胚形成前细胞化。在果蝇中,前后体轴被**母体基因**(maternal gene,在母细胞中表达的基因,其产物转移到卵细胞)特异化,它们在合胞体胚胎中建立了前部、后部和末端组织中心。这些基因编码调节因子建立了相反的形态发生梯度,以浓度依赖的方式激活下游的**合子基因**(zygotic,在发育中胚胎表达的基因,而不是在周围母细胞中表达的基因)。连续的级联激活不同类型的合子基因,将合胞体分成为由基因表达区域定义的条状区。这些形成了体节的基础,成为由后期作用的基因模式。

前后部中的母体基因 控制果蝇前后组织有四个基本的基因:它们是特异化前方结构的 *bicoid* 和 *hunchback*,和特异化后方结构的 *nanos* 和 *caudal*。在每种情况中,mRNA 都是在滋养细胞中合成,它位于卵的前部。

bicoid 的 3′端非翻译区(UTR)通过 *exuperantia* 和 *swallow* 基因产物附着于前方细胞骨架上。*bicoid* mRNA 的翻译开始被其非常短的多聚腺嘌呤尾巴所阻止,但受精后,*cortex*,*grauzone* 和 *staufen* 基因产物促进进一步多聚腺嘌呤化,使 Bicoid 蛋白合成。*nanos* mRNA 通过 *oskar*,*staufen*,*tudor*,*vasa* 和 *valois* 基因产物,转移到卵细胞的后部,它的 3′UTR 与后部细胞骨架相结合。*hunchback* 和 *caudal* 的 mRNA 不定位,在整个卵细胞中分散分布。

Bicoid 是含同源异型结构域的蛋白,与 DNA 和 RNA 结合。它在合子细胞核中激活 *hunchback* 基因的转录,抑制母体 *caudal* mRNA 在卵细胞前部的翻译。相反的是,Nanos 蛋白抑制母体 *hunchback* mRNA 在卵细胞后部的翻译(它不是转录因子)。Pumilio 蛋白与 *hunchback* mRNA 结合,为 Nanos 提供进入位置。四个基因产物的相互作用产生了一个 Bicoid 和 Hunchback 的前-后梯度及 Nanos 和 Caudal 的后-前梯度。

末端的母体基因 末端的基因特异化**顶节**(acron)和**尾节**(telson),这样的结构形成了前后轴的末端。这个群体中的关键因素是 *troso* 和 *torsolike* 基因。Torso 是酪氨酸激酶类型的一个跨膜受体。*troso* 的 mRNA 在卵细胞中广泛分布,它编码的受体在整个细胞质膜上分布。末端特异化是被配体 Torsolike 控制,Torsolike 只从前后的卵泡细胞中分泌。Torso 被其配体激活,诱导一个涉及 Ras、Raf 和 MAP 激酶的信号通路(见信号传导),在一个诱导裂缺(gap)基因 *huckebein* 和 *tailless* 转录的未知转录因子激活时达到顶峰。这些基因编码激活特异化末端结构基因的转录因子。顶节特异基因也需要 Bicoid 蛋白。持续激活的 Torso 突变体上位于前后部基因-表型是"超末端"的,有大的顶节和尾节,但它们中间没有身体。

前后合子基因 母体基因 *bicoid*,*hunchback* 和 *caudal* 编码控制合子基因表达的转录因子,涉及体节形成的规划。分成五类合子基因:

裂缺基因(gap genes)将胚胎分成对应副节的主要区域,缺失功能突变导致这些副节在幼虫中丢失(**副体节**包含一个体节的前面部分和邻近体节的后面部分;胚胎期果蝇身体的划分是通过基因表达区域分成副节,虽可识别的形态结构是体节)。

初级配对规则基因(primary pair rule genes) 为裂缺蛋白调节,在不同的副节中表达,这些结构在功能缺失突变中缺失。

次级配对规则基因(secondary pair rule genes) 与上面一样,但比初级配对规则基因作用晚,需要初级配对规则基因和裂缺蛋白来建立它们的表达区域。配对规则基因一起确定胚胎中的副节。

体节极性基因(segment polarity genes)　是在连续副节和每个相同的相对位置表达。它们负责副节中细胞类型的特异化,功能缺失突变引起相应副节的缺失,取而代之的是呈镜像的剩余部分。

同源异型选择基因(homeotic selector genes)　在包含一个或几个副节的特定区域中表达,特异化体节的特征,也就是使每个体节将发育成怎样的结构。

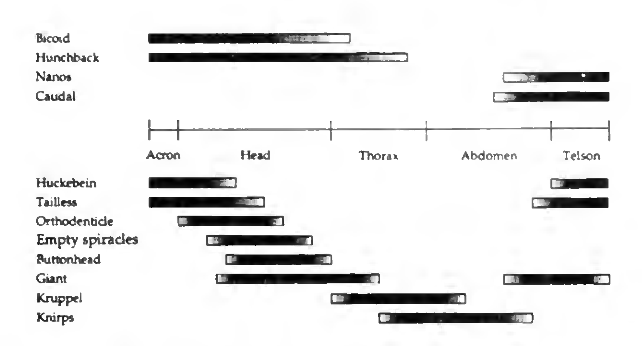
裂缺基因和配对规则基因因此解译了定义轴的母体信息,将其转换成重复的体节结构,而体节极性基因和同源异型基因赋予体节位置值。裂缺基因和配对规则基因在下面将更详细讨论,体节极性基因和同源异型选择基因的讨论见文框6.8。

裂缺基因　裂缺基因开始在广泛区域中表达,当Hunchback蛋白梯度稳定后定位。Bicoid和Hunchback在胚胎前部激活 *giant* 和 *Krüppel*,抑制 *knirps*。Caudal蛋白负责胚胎后部的 *giant* 和 *knirps* 的转录。三个头部特异基因 *orthodenticle*, *empty spiracles* 和 *buttonhead* 只在有高浓度Bicoid蛋白的地方表达。*tailless* 和 *huckebein* 基因受到胚胎末端Torso信号激活的转录因子调节。产生的裂缺基因表达模式在下面可以看到。裂缺基因的产物本身都是转录调节因子,它们有两个功能——相互反馈和调节,以及调节下游配对规则基因的表达。这样的调节看来基本上都是负调节。一旦mRNA翻译,蛋白分向两侧分散,抑制邻近裂缺基因的表达,这样建立了明显的转录边界。

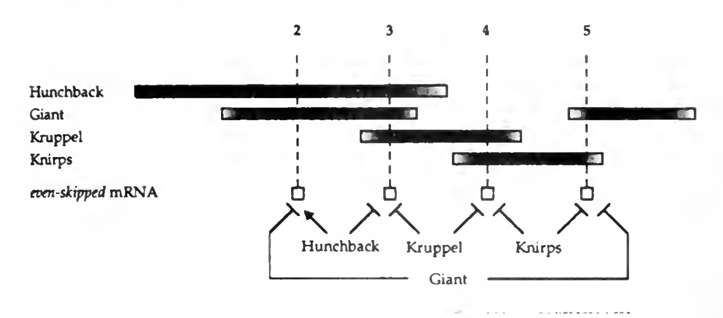
配对规则基因　配对规则基因在中胚层细胞化过程中表达,mRNA沿前后轴呈斑纹分布。每个斑纹被所在位置裂缺基因产物单独调节。不同配对规则基因对不同裂缺蛋白浓度反应,所以在重复模式中,邻近斑纹细胞表达不同组合的配对规则基因产物。有三个基本的配对规则基因,*even-skipped*, *hairy* 和 *runt*,它们的表达对分节模式的建立很重要。基本配对规则基因的表达模式通过交互调节稳定,帮助特异化次级配对规则基因产生的模式,这包括 *fushi tarazu* 和 *odd-paired*。初级配对规则基因 *even-skipped*(*eve*)的调控研究得最清楚,它主要是负调控。eve mRNA在大多数裂缺蛋白低的地方表达,也就是在裂缺蛋白基因区域的边界。如下所示,每个 *eve* 表达斑纹受几个裂缺基因的控制,这反映了 *eve* 启动子的调节结构,有分离的调节元件控制斑纹4～6、斑纹1、2、3、7。斑纹2的调控元件被细致地作图,启动子的这个区域包含Bicoid、Hunchback、Giant和Kruppel的结合位点。正调节因子(Bicoid和Hunchback)的结合位点经常与负调节因子(Giant和Kruppel)的重叠。表明不同调节因子的竞争是斑纹定位的重要机制。

背腹极性　果蝇的背腹极性是在细胞化后通过核Dorsal蛋白的梯度建立的。Dorsal在果蝇中普遍合成的,但它特异性转移到胚胎腹侧的细胞核中。在胚胎背侧,Dorsal与第二个蛋白Cactus相联系,Cactus掩蔽Dorsal核定位信号。Cactus的降解是被磷酸化诱导的,磷酸化是通过从胚胎腹侧发出的信号传导反应。Dorsal是与脊椎动物转录因子NF-κB同源,而Cactus与抑制蛋白I-κB同源;激活的信号途径也是保守的。背腹极性的途径从卵母细胞本身开始,通过 *gurken* 和 *cornichon* 基因产物使卵泡细胞背侧化。Gurken是与哺乳动物表皮生长因子(EFG)同源,在卵泡细胞膜上激活受体Torpedo(与EGF受体同源)。Torpedo受体的激活诱导抑制 *nudel*, *windbeutel* 和 *pipe* 基因的信号途径。在胚胎的腹部,这些基因是活性状态的,它们的产物形成膜复合物,其功能是激活由胚胎分泌的,*easter*, *snake* 和 *gastrulation defective* 编码的三个丝氨酸蛋白酶。Gastrulation defective的激活起始蛋白酶级联反应:Gastrulation defective剪切Snake,Snake剪切Easter,Easter再剪切另一个背侧基因 *spatzle* 产物,*spatzle* 产物是Toll受体的配体。Toll在卵细胞膜上广泛分布,但因为蛋白酶级联反应只在腹侧激活Spatzle,Toll信号也是腹侧特异的。Toll激活涉及 *tube* 基因的信号传导途径中的蛋白酪氨酸激酶Pelle。Pelle是磷酸化Cactus的酶,诱导它的降解,允许Dorsal蛋白进入腹侧细胞核。在细胞核中,Dorsal同时作为转录激活因子和抑制因子。它激活 *rhomboid*、*snail*

和 *twist* 基因，而抑制背侧化基因 *zerknullt*、*tolloid* 和 *decapentaplegic*。



在 12 核分裂后多核体果蝇胚胎中关键的母系卵极性基因（上）与合子裂缺基因（下）的产物分布。



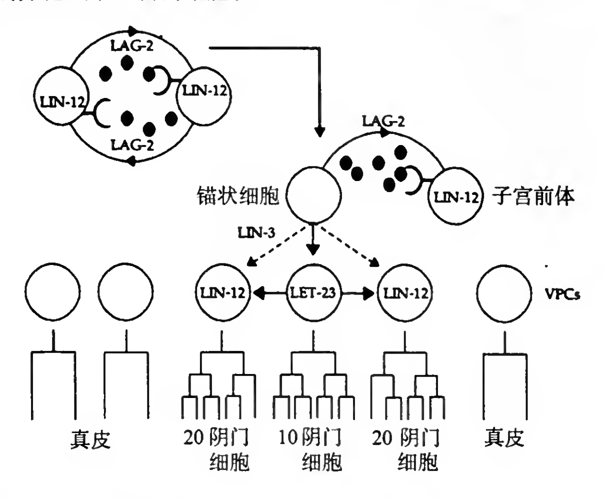
合子基因产物对 *even-skipped* 转录的调节。在裂缺蛋白质水平最低处定位了 *eve* mRNA 斑纹，除了由 Bicoid 和 Hunchbake 正向调控的斑纹 2。

文框 6.6 线虫（*C. elegans*）的阴门特异化

阴门系 线虫的阴门（vulva）是腹部真皮（动物的外层，相当与哺乳动物的真皮）中的开口，通过阴门卵细胞从上面的子宫远处移出。阴门由 22 个细胞组成，从三个真皮细胞经过三次细胞分裂，而其他真皮细胞只分裂一次释放出产生真皮细胞。真皮细胞通过性腺上称为**锚状细胞**（anchor-cell）的一个单细胞，诱导形成阴门。如果锚状细胞被破坏，所有的腹部真皮细胞只分裂一次，形成真皮。如果所有的性腺细胞除了锚状细胞都被破坏，阴门还会形成。如果锚状细胞被机械去除，邻近的真皮细胞能够对此反应，形成阴门，但只有六个真皮细胞有能力这样做。

阴门前体细胞表型特异化 涉及阴门特异化的基因通过无阴门和多阴门表型的突变发生进行筛选鉴别，证明在简单系统中确定发育途径的遗传力量。阴门模型提供了几种类型细胞相互作用的例子。起先，两个相当的细胞有能力变成锚状细胞，接受它们之间的信号，涉及信号分子 LAG-2 和受体 LIN-12，结果在侧面抑制其中一个细胞。如果通过 *lag-2* 或 *lin-2* 任一突变消除信号，两个细胞都成为锚状细胞。在持续信号显型功能获得 *lin-12* 等位基因突变体中，两个细胞都变成子宫组成的前体细胞。

一旦锚状细胞特异化，它诱导下面的下胚层细胞形成阴门。中心细胞成为中心阴门细胞，两侧细胞为侧部阴门细胞。特别有意思的是编码信号途径基因与哺乳动物的 RTK-Ras-Raf 途径同源(见信号传导)。由 *lin* -*3* 基因编码的配体是生长因子 EGF 家族的成员，是锚状细胞所需的。其他成分由 *let* -*23*，*sem* -*5*，*let* -*60* 和 *lin* -*45* 编码，是下胚层细胞反应和所需，对应于受体，接头蛋白，Ras 和 Raf。MEK 和 MAP 激酶的同源蛋白也已经被鉴别。LIN - 3 可能是引起下面的阴门前体细胞和侧部细胞不同反应的形态生成因子。中心细胞也分泌额外的信号，诱导两侧细胞 *lin* - *12* 的表达，阻止它们分化成中心阴门细胞。



线虫(*C. elegans*)阴门特异化总结。侧抑制建立了锚细胞，锚细胞再诱导下面的下胚层细胞分化为阴门细胞，初级和次级细胞系可通过 LIN - 3 的浓度梯度或通过来自中心阴门细胞的信号而特化。

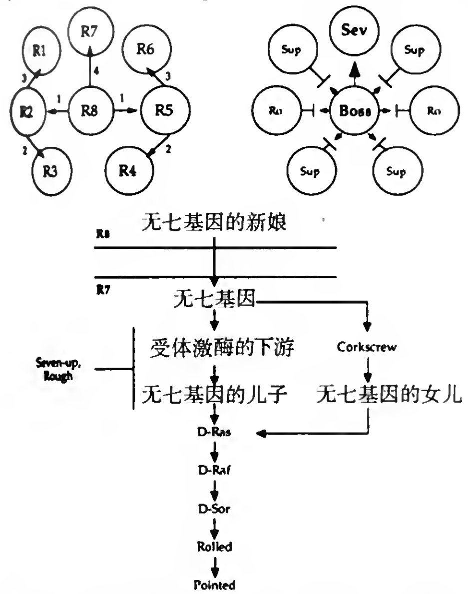
文框 6.7 果蝇眼的发育

光受体细胞：果蝇的复眼包括大约 800 个功能单位，称为 **omatidia**，每个由 20 个细胞组成，其中八个是光受体。眼细胞分化涉及在单细胞水平上一系列结构诱导性级联调节，晶状体代表了缺乏状态(非诱导)。眼起源于一层细胞，omatidia 分化通过程序性的细胞 - 细胞相互作用而发生。决定第一个细胞的是中心 R8 细胞。其机制还不清楚，但可能涉及信号分子，如在跨越眼原基的形态发生沟中 Hedgehog 和 Decapentaplegic。从 R8 发散的信号诱导沿前后轴邻近细胞分化成为功能相等的 R2 和 R5 细胞。这些细胞再将信号传递到附近两侧未分化细胞，形成 R1，R3，R4 和 R6 细胞，它们也是功能相当的。最后是 R7 光受体细胞的特异化，其机制已经研究得较完善。所有其他细胞成为晶状体。

R7 光受体的特异化　　通过异常 R7 细胞发育表型发现了很多基因。无七基因[*sevenless*(*sev*)]和无七基因的新娘[*bride of sevenless*(*boss*)]似乎是在途径开始时发挥功能的，镶嵌分析表明 *sev* 在将来 R7 细胞所需要而 *boss* 在诱导 R8 细胞中需要。Boss 是激活 Sevenless 受体酪氨酸激酶的信号分子，起始 R7 细胞中的信号传递反应。这个信号途径中的很多成分已经被鉴别——Drk(downstream of receptor kinase，下游受体激酶)是有 SH2 和 SH3 结构域的接头蛋白，无七基因的儿子[Sos(son of sevenless)]是鸟苷酸交换因子，作用于 Ras。Ras，Raf 和 MAP 激酶同源蛋白也被发现(尽管

这些基因突变是多效应的和致死的，因为同样的蛋白还调节果蝇中其他信号途径）。看来途径中还有其他不同的支路，涉及蛋白酪氨酸磷酸酶 Corkscrew 和它的底物无七基因的女儿[Dos(daughter of sevenless)]，Dos 也可以激活 Ras。

R7 的决定反映了存在一种由 R8 细胞显示的可对 Boss 信号反应的受体和信号途径。然而所有的光受体细胞最初合成 Sevenless，这就意味着在其他细胞中须存在抑制途径的机制。一个称为 Sevenup 的转录因子可能在 R1，R3，R4 和 R6 细胞中调节这种功能，因为功能缺失突变导致这四个细胞分化成类 R7 细胞。在 R2 和 R5 细胞中表达 Rough 转录因子是 R3 和 R4 细胞中 *sup* 表达所需，可能类似的机制诱导 R1 和 R6 中 *sup* 的表达，在 R2 和 R5 本身抑制 Boss－Sev 信号传导途径。



果蝇眼发育。（左图）特化光受体细胞命运的诱导反应的次序。（右图）R7 光受体的特异性受其他细胞中阻断 Boss-Sev 信号转导的调节子控制。（上图）Boss-Sev 信号转导途径到 R7 的分化（见信号转导）。

文框 6.8 果蝇中分节化和体节特性

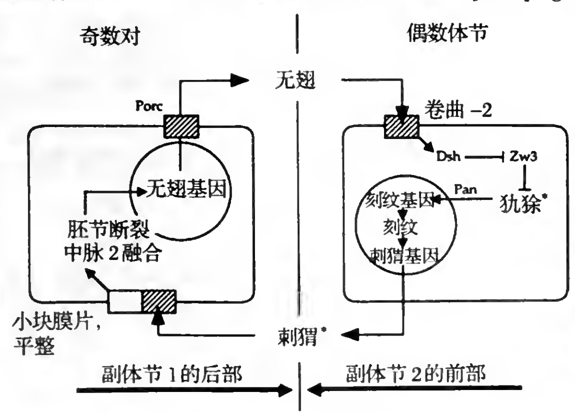
体节极性基因 母体基因，前后轴的合子裂缺基因和配对规则基因都编码基因表达调节因子（不是转录就是翻译的），因为发育到这点时合胞体已经发生，而调节因子自由扩散在细胞核内相互作用（见文框 6.5）。在配对规则基因表达时发生细胞化，因此下游过程发生在多细胞环境中进行。体节极性基因编码信号蛋白和转录因子。

体节极性基因有两个功能——它们通过接受信号维持和加强副节的分节模式和它们在每个副节中建立单个细胞的归宿。第一种加工在副节两侧细胞中研究得较清楚。起先，表达 *engrailed* 的细胞在 *even-skipped* 或 *fushitarazu* 的控制下，从每个副节前部边缘产生。表达 *wingless* 的细胞从每个副体节不表达 *eve* 和 *ftz* 的后部产生，因此 *wingless* 可能被 Odd-paired 或其他配对规则蛋白调节。一旦建立，*wingless* 和 *engrailed* 的表达通过接受下面阐述的信号途径维持。分泌的 Wingless 激活 Frizzled 受体和细胞内信号，结果激活转录因子 Armadillo，Armadillo 激活 *engrailed* 基因的转录。Engrailed 诱导 *hedgehog* 的转录。它编码一个与邻近细胞上 Patched 受体作用的信号分子，

将信号蛋白 Smoothened 从抑制状态释放出来。通过胚节阻断而 Smoothened 的信号诱导 *wingless* 基因的转录,使细胞分泌 Wingless 蛋白。

第二个过程了解较少。在不同的配对规则基因的控制下,体节极性基因在每个副体节中特定细胞中表达。这样它们可以调节导致每个细胞分化成区域正确细胞类型的基因。然而,上述的 Wingless 和 Engrailed 细胞看来在这一过程中起主要作用,表明 Wingless 和 Engrailed 与建立副节分界一样,在副节中传播分级信号,特化单个细胞的命运。

同源异型选择基因　同源异型选择基因提供体节位置信息,也就是它们控制体节的单个发育,从共同的细胞类型群产生区域专一性结构。因此这些基因的突变有将身体一部分转换成类似另一部分的效果,称为**同源异型转化**(homeotic transformation)。大多数的果蝇同源异型基因在染色体 3 的两个基因簇中发现,称为**触角足复合体**[antennapedia, ANT-C,包含基因有唇(*labial*),触角足(*antennapeia*),性梳减少(*sex combs reduced*),畸形(*deformed*)和吻足(*proboscipedia*),特异化头和胸的体节]和**双胸复合体**[bithorax complex, BX－C,包含基因有超双胸(*Ultrabithorax*),腹部 A 和腹部 B(*abdominal A* 和 *Abdominal B*),特异化腹部体节]。这些一起组成**同源异型复合体,HOM-C**。同源异型基因在胚胎的特定区域表达,在有些情况中,对应于特定的体节或副节,在其他跨越几个副体节(这些表达参数在文框 6.9 中显示了与哺乳动物 *Hox-B* 基因簇基因的比较)。值得注意的是,沿染色体基因的顺序通过胚胎表达图谱的空间区域来概括,3′端基因在大多数前部区域表达(见文框 6.9)。同源异型基因表达由裂缺基因和配对规则基因产物的活性来起始。象体节极性基因,同源异型基因通过交互调节维持它们的表达模式。因此同源异型突变的效果不仅是对突变的基本反应,而且是对其他同源异型基因区域的影响。再者,交叉调节在自然界中主要是负调节,每个异型基因被在更后面区域表达的基因所抑制。失去功能的同源异型突变因此增加了前面区域的基因,结果使后面体节变成前面一个。同源异型基因转录活性通过染色质结构的修饰而稳定。无活性的同源异型基因隐蔽在被抑制的染色质中,这样的染色质被多梳(Polycomb)家族的蛋白所稳定。相反的,三胸(Trithorax)家族蛋白似乎是维持活性染色质区域(见染色质)。同源异型基因编码的转录调节因子包含保守的 DNA 结合结构域,称为同源异型结构域(见核酸结合蛋白)。它们在体外似乎有重叠的 DNA 结合专一性,在体内被认为与辅助因子一起作用确定真正的 DNA 结合位置。一些可能的辅助因子已经被发现,包括外小齿(*Extradenticle*)和 *teashirt* 基因的产物。同源异型结构域蛋白调节的很多下游靶基因也被发现。*salm* 和 *distal-less* 基因分别是眼和腿发育所需的,它们本身编码转录因子,而 *decapentaplegic* 编码涉及腿命运特异化的信号分子。

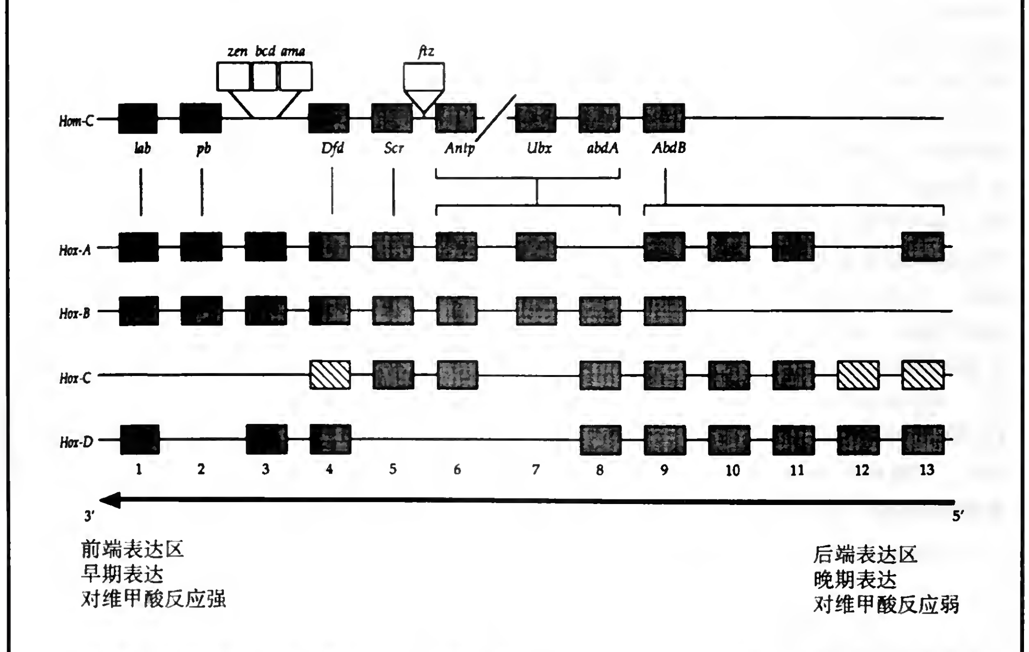


果蝇中维持副节分界的信号途径的总结。*engrailed* 基因开始由 Even-skipped 或 Fushi tarazu 活化,而相邻细胞中的 wingless 由另外的配对规则蛋白所活化,可能是 Odd-paired。配对规则基因的表达是瞬时的,但 wingless 与 Engrailed 一旦活化就可以通过交互信号彼此维持。

* 果蝇早期发育中的一些突变体及相应基因常被人们用一些形象的词汇来描叙,如两头尖(biocoid),驼背(hunchback)等,图中的犰狳(Armadillo)和刺猬(hodgehog)也是用来描叙突变体表型特征的。——译者注

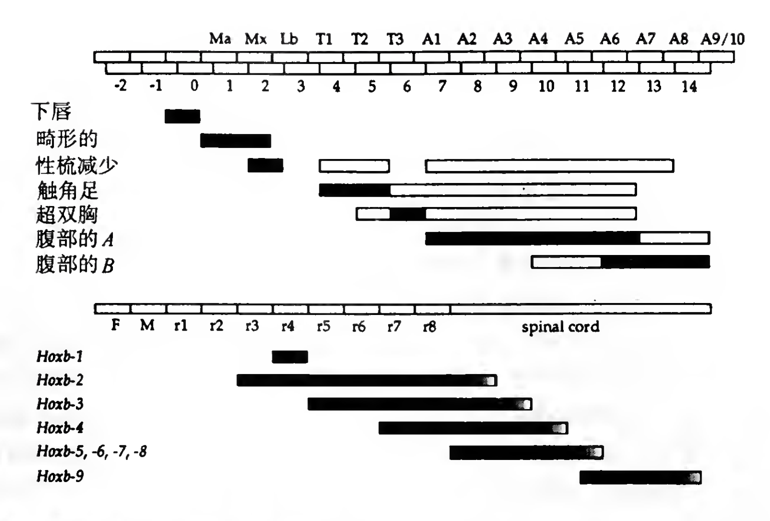
文框 6.9 脊椎动物 *Hox* 基因

脊椎动物基因组包含了四个拷贝的果蝇同源异型复合物,称为 *Hox-A*,*Hox-B*,*Hox-C* 和 *Hox-D*。尽管脊椎动物和果蝇有明显的差异,这种类似性是引人注目的。同样类型的同源异型盒基因出现,根据同源异型盒结构分成 13 个相关组(cognate groups)(**共生同源亚群**,paralogous subgroups)。进而基因沿染色体以或多或少相同的顺序排列,以类似的方式表达,3′端基因在大多数前部结构域表达,而 5′端基因在大多数后部分结构域表达。因为果蝇和脊椎动物的差异,果蝇 HOM-C 分成两个亚复合物,而脊椎动物基因簇进行 5′末端扩增,整个复制后产生四个复合物。每个四复合物有单个的缺失,这在物种间存在差异。如在上图中代表 *Hox-C* 基因的有阴影的框盒在人类中存在,但在小鼠中没有。



果蝇和脊椎动物 Hox 基因簇的结构。果蝇基因簇被包括 *bicoid* 的非 *Hox* 基因打断。原始的脊椎动物基因簇在复制前进行 5′扩增,几个基因被丢失。

根据结构(见上图)和表达分布(见下图)这二者,果蝇和脊椎动物同源异型盒包含基因的类似性是保守功能的有力证据。用克隆的人类 *HOX* 基因恢复果蝇同源异型突变和定向破坏小鼠的 *Hox* 基因确实产生部分同源异型转化(参见基因敲除)证实上述的保守功能。如 *Hoxc-8* 基因的缺失,结果腰椎部分转化成有更多前部特征的椎骨(在这种情况下是有肋骨的完善的胸锥)。副基因



果蝇 HOM-C 复合物和小鼠 *Hox-B* 基因簇在神经系统中的表达。

座 *Hox* 基因看来有重叠但功能不完全一样，在特定情况下相互协作。在副基因座亚群 3 中单个基因的敲除，引起颈结构不同类型的分布，但在同一只小鼠中组合时，可以观察到严重的缺陷，包括椎骨的缺失。

文框 6.10 脊椎动物肢的发育

脊椎动物肢的起源和发育 脊椎动物肢从肢区(limb fields)发育得到，肢区通过前后方向 *Hox* 基因确定的体轴特定位置产生(见文框 6.9)．侧板中胚层(形成骨骼元件)和体节(形成肌肉元件)的生长形成称为肢芽的突出。生长可以被中期肾分泌的 FGF－8 所诱导。肢发育的一个重要事件是**顶外胚层嵴**(apical ectodermal redge, AER)的形成。这是外胚层升起的顶部，在中胚层下维持生长，并在轴特异化中起重要的调节作用。只有外胚层在胚胎的背/腹边缘才有能力形成 AER，这反映了称为主缘膜(radical fringe)在背外胚层的合成。主缘膜可能涉及接受邻近腹侧表达 *en* -2 细胞的信号，*en* -2 是果蝇 *engrailed* 基因的同源基因。边缘细胞被诱导升高并合成 FGF－8，使它们支持下面中胚层的生长。随着肢芽的延伸，一个不连续的大约 200μm 的**前进区域**可以在 AER 下面发现，在这个区域中细胞继续生长。在前进区域后面，细胞分化成沿远近方向(手臂到手指)、前后方向(拇指到手指)和背腹方向(指节到手掌)轴，对应它们位置正确的结构。前进区域的大小反映了 FGF-8 信号的范围。

前肢或者后肢？ 分化成前肢或后肢结构的决定反映了前后 *Hox* 编码的中胚层肢的成分中的位置信息。在前和后肢中有不同的 *Hox* 基因组合和其他转录因子(如 Tbx4 和 Tbx5 分别在小鼠的前肢和后肢表达)。这些因子控制了对形态发生指令的反应，如生长和分化信号，使细胞在每个肢中不同表现，形成肢特异性结构。

远近方向轴 沿远近方向轴的区域专一性分化(如手臂、肱部、尺骨、掌骨或手指)反映了细胞保留在前进区域中时间的长度。异时嫁接(一个肢芽的前进区域被嫁接到另一个 AER 下面形成不

同的发育阶段)显示了肢的发育与提供的前进区域所携带的指令一致。离开前进区域的老细胞持续表达 *Hox-A* 和 *Hox-D* 基因簇的末梢基因,副基因座 *Hox* 基因的混合敲除引起特定远近结构的缺失。这样看来开始分化时细胞的年龄多半通过它进行分裂的次数显示,可能以某些方式调节 *Hox* 基因表达,并输入远近位置信息到分化细胞。

背腹方向轴 肢的背侧是通过信号分子 Wnt-7a 特异化的,Wnt-7a 在背外胚层表达。缺乏 *wnt-7a* 基因的小鼠有腹肢,缺乏后部结构因为 *Wnt-7a* 是前后轴特异化所需(见下文)。一个转录因子 Lmx1 被发现是背特异化所需,它的基因被 Wnt-7a 信号途径诱导。

前后方向轴 前后方向轴是通过后部的**极性活性区域**(zone of polarizing activity,**ZPA**)特异化的,ZPA 在 AER 形成时从肢区形成的。ZPA 作为一个组织者,当嫁接到肢的不同位置,诱导第二个轴。从 ZPA 散发的信号诱导末梢 *Hox-D* 基因的浓度嵌套式表达分布,敲除破坏和异位表达实验表明特定 *Hox-D* 基因组合决定了前后位置值,如形成趾的类型。*Hox-D* 表达分布的建立是以一涉及 ZPA 豪猪(Sonic hedgehog,Shh)蛋白 Shh 的复杂的方式,Shh 与其他在 AER 中合成的信号分子 FGF－4 和 BMP－2 协同作用。

从早期 AER 分泌的 FGF－8 和在背外胚层表达的 Wnt－7a 起始 *shh* 的表达。*shh* 的表达限制在肢芽的后部,而前部边缘表达 *hoxb*－*8* 。Shh 诱导在 AER 后部的 *fgf*－*4* 表达,两个蛋白通过自动调节环路相互维持。维甲酸诱导 *shh* 表达,模拟 ZPA,可能也涉及诱导 *fgf*－*4* 的表达。因此 Shh 的表达和功能是依赖于涉及所有肢轴特异化的分子。

果蝇和脊椎动物中的位置信息 肢的模型表明 *Hox* 基因在特异化位置信息方面与果蝇中同源异型基因显示了类似的作用。在两种生物中,*Hox* 基因的缺失造成同源异型转化,尽管在脊椎动物中这种情况通过有些成分的过剩而复杂化。果蝇中的体节极性基因是维持细胞边界和建立细胞命运和功能所需,不仅在体节特异化,而且在其他结构中,包括腿和翅。在轴特异化中特别重要的是 Hedgehog、Wingliess 和 Engrailed 蛋白的相互作用。在脊椎动物的肢和其他区域中,都涉及同样这三个分子。其次因为存在着重复,故果蝇和脊椎动物有差异,在小鼠中有三个 Hedgehog 相关分子、多于 10 个 Wingless 同源蛋白和两个 Engrailed 相关转录因子。在肢芽中,Shh、Wnt－7a (Wingless 相关蛋白)和 Engrailed－2 在体轴特异化起重要作用,并作用于上游的 *Hox* 基因。值得注意的是很多果蝇区域特异化网络中的成分在脊椎动物中反复使用,而它们早期发育是基本不同的。

(徐 沁 译)

第7章
DNA 甲基化及外遗传调节

基本概念和定义

- 核酸和蛋白质在合成或合成后可通过某些特殊化学基团的加入进行修饰。RNA和蛋白质经常发生修饰(见第27章RNA的加工,和第22章蛋白质),而DNA分子的修饰是有限的。一些病毒改变其基因组中特定的核苷酸残基,来保护DNA免受核酸酶的降解或促进病毒包装到衣壳中(见表7.1)。**DNA 甲基化**(DNA methylation)是一个主要的例外,它是在原核生物和高等真核生物中经常发生的现象,在DNA识别和功能中起多方面的作用。
- DNA的甲基化和病毒基因组的修饰都涉及DNA碱基的共价修饰,但并不改变碱基配对的专一性,也并不改变DNA所携带的遗传信息(参考突变)。甲基化可以影响DNA的功能,例如改变DNA与转录因子以及其他一些蛋白因子的相互作用方式。因此DNA的甲基化状态可以调节基因的表达并直接影响生物的表现型。
- 将甲基加入到DNA中去的酶称为**DNA 甲基转移酶**(或称为**DNA 甲基化酶**),它利用S-腺苷酰甲硫氨酸作为甲基的供体。体内两种甲基化酶的活性可被区分:一种称为从头合成甲基化的酶,它将甲基加到DNA的特定位点,启动一种DNA甲基化的模式。而维持甲基化酶将甲基加到一条链已甲基化的DNA(**半甲基化DNA**)上。这样通过连续的复制使甲基化的模式保持不变。维持甲基化酶的靶位点通常是二重对称的,因此同一种酶能甲基化两条子链的新生链。
- DNA携带着两种形式的信息:由核苷酸序列决定的**遗传信息**和由它的结构决定的**外遗传信息**。外遗传信息是不处于DNA自身的核苷酸序列中的可影响DNA活性的任何可遗传的性质(也就是说,最终将影响器官的表现型)。DNA的甲基化是外遗传的一种:它是可遗传的(由维持甲基化酶催化),而且能够调控基因的表达。因此甲基化的表现型被称为**后生型**(epigenotype),而甲基化的改变称为**后生突变**(epimutation)。外遗传信息的其他形式还有染色质的结构(见染色质)和DNA分子的构象和拓扑特征(见核酸结构)。

7.1 原核细胞的DNA甲基化

限制修饰系统的甲基化 许多细菌能编码以序列特异方式剪切DNA的内切核酸酶。

表 7.1　病毒基因组中 DNA 核苷酸的修饰

病毒	核苷酸的修饰
FV3	约 20%胞苷酸甲基化为 5-甲基胞苷酸
φW-14	胸苷生成 α-1,4-丁二胺脱氧胸苷
PBS1	胸苷形成脱氧尿苷
SPO1	胸苷形成 5-羟基尿苷酸
T-偶数家族	胞苷酸生成羟甲基胞苷酸

注:FV3 是一种真核病毒,它能编码自身 DNA 的胞嘧啶甲基转移酶。而其余四种都是噬菌体病毒。

这些酶能在外源性 DNA(例如噬菌体基因组)进入细胞时切断外源 DNA 从而建立一种保护机制。因此敏感性噬菌体在转染编码 DNA 内切核酸酶的宿主菌时效率较低,而在不编码 DNA 内切核酸酶的宿主菌中的转染效率有很大的提高。由于宿主菌能通过它所编码的 DNA 内切核酸酶限制相应的噬菌体在胞内的繁殖,因此这类酶被称为**限制性内切酶**(restriction endonucleases)。

宿主菌通常通过在内切核酸酶识别位点中特定碱基的甲基化来保护自身 DNA 免受内切核酸酶的酶解。因而每一种 DNA 内切酶都有一个**相关甲基化酶**(cognate methylase),这种酶可能是一个独立的酶,也可能是一个通用的全酶的一部分。这样就组成了一种**限制修饰系统**。这在细菌中普遍存在,目前已分离出数百种限制性内切酶和甲基化酶。它们在体外 DNA 实验操作中已被广泛采用(见第 24 章重组 DNA)。许多 DNA 噬菌体也进化产生了多种对抗限制的策略,如改变自身基因组 DNA 或合成限制性内切酶的抑制剂。很多噬菌体在相应的宿主菌内并不受限制性内切核酸酶的作用。很多病毒也缺少其宿主的限制靶位点。

Dam 甲基化　　在大肠杆菌中,**DNA 腺嘌呤甲基化酶(Dam)**通常催化 GATC 序列中腺嘌呤残基的第六位氮原子(N^6)甲基化。该序列表现出二重对称,被作用时两条链上的腺嘌呤残基都发生甲基化。这种修饰机制的主要作用是使细胞识别 DNA 复制后的母链和暂时还未甲基化的子链,这促进了校正复制错误产生错配的 DNA 修复系统,即**复制后错配修复**(参阅)的作用。子链的非甲基化作用将指导修复酶从子链上切除错误的核苷酸而不是切除母链上正确的核苷酸。

DNA 腺嘌呤甲基化酶在 DNA 的复制中也起着直接的作用。在大肠杆菌复制起始区 *oriC* 中有 14 个腺嘌呤甲基化位点,编码必需的复制起始蛋白(参见**复制起始**)的 *dnaA* 基因的启动子上也发现了另一个甲基化位点。在复制开始后半甲基化的起点可能由于与细胞膜中的成分相互作用而无法再次启动 DNA 的复制。子链 *oriC* 起始区和 *dnaA* 基因启动子上的甲基化位点通常比一般 Dam 位点中的甲基化更晚一些,并可能受**细菌细胞周期因子**(参阅)的直接调控。*DnaA* 基因启动子中的 GATC 序列可能作为一种转录沉默子而阻止转录的再起始。类似的,多种细菌转位子中的 GATC 序列可能通过阻断转位酶基因的转录而阻断基因的转位,或者在转录后阻断转位酶与 DNA 的结合而发挥作用。在两种情况中,易位被限制在较短的时间中,从而确保胞内存在两个基因组以利重组介导的损伤修复。有趣的是,DNA 甲基化在真核基因的转位中也起着重要作用,这将在下文中讨论(见第 13 章可移动的遗传因子)。

Dcm 甲基化 在大肠杆菌中,CCWGG 序列中的胞嘧啶残基在 **DNA 胞嘧啶甲基化酶**(**Dcm**)的催化下转变成 5-甲基胞嘧啶。这种甲基化的作用目前还不清楚,它可能保护基因组免受限制性内切酶 *Eco*RII 的酶解。5-甲基胞嘧啶容易自发脱氨而形成胸腺嘧啶,会导致 G-T 错配。*Vsr* 基因编码的 DNA 修复系统通常能纠正这一错误(参见极短补丁错配修复)。值得注意的是,*vsr* 基因和 *dcm* 基因位于同一个操纵子上,它们的读框有 6 个密码子相互重叠。

7.2 真核细胞的 DNA 甲基化

真核细胞中胞嘧啶甲基化的模式 在真核细胞 DNA 中发生修饰的碱基一般只有 5-甲基胞嘧啶(5-meC),并可能是仅有的程序化修饰。5-甲基胞嘧啶的丰度在不同的物种间差异较大,通常在真菌和无脊椎动物中含量较少(例如在酿酒酵母和果蝇中 5-甲基胞嘧啶检测不到),在脊椎动物中有一定的含量(高至 10% 的胞嘧啶残基会发生甲基化),而在很多植物中含量最高(最高可有 30% 的胞嘧啶残基发生甲基化)。

在脊椎动物细胞中,胞嘧啶残基甲基化通常发生在 5′-CG-3′二核苷酸基序中,而在植物细胞中 5′-CG-3′和 5′-CNG-3′这两种二重对称的基序都可被甲基化。目前分离得到的真核细胞 DNA 甲基化酶多是维持甲基化酶,很多真核基因组中观察到的 DNA 甲基化是组成型的,甲基化的维持对其增殖已足够。下面还将讨论到,在动物和植物发育过程中观察到 DNA 甲基化状态的全局性变化证实了从头合成甲基化酶和去甲基化酶的存在,尽管这两类酶的性质目前未被描绘。另外,在非对称序列上也可能发生胞嘧啶甲基化,该甲基化可能受 DNA 复制的调控(参见深染甲基化岛),这些位点的甲基化修饰也需要从头合成甲基化酶的活性。

胞嘧啶的甲基化,突变和基因组进化 在脊椎动物细胞 DNA 中,由于 5-甲基胞嘧啶的高突变率,CG 出现的频率只有按核苷酸组成计算出的频率的 20%。脱氨基作用(参阅)是 DNA 自发损伤的一种普遍形式,未发生修饰的胞嘧啶的脱氨基产物是尿嘧啶,它可以通过碱基切除修复(参阅)从 DNA 上有效切除。而 5-甲基胞嘧啶的脱氨基产物是胸腺嘧啶是一种合法的 DNA 碱基,在细菌中存在一种特殊的修复系统来纠正由此产生的 T:G 错配。但这种修复作用是泄漏型的,因而导致胞嘧啶随着进化历程含量下降。在人类点突变引起的疾病中 C→T 的转换占大多数。

CpG 岛 通过分析 5-甲基胞嘧啶在真核细胞基因组中的分布可以发现 **CpG 岛**的存在(或称为 **HTF 岛**[1])这些区域是非甲基化的,通常为 1～2kbp 大小,并且绝大多数结合在一些基因的 5′末端。这些序列富含 GC,并且有大量的 CpG 二核苷酸基序。这表明甲基化的去除防止了胞嘧啶残基的缺失。CpG 岛的这些性质可以用于诊断或鉴别一些大的基因组克隆中的潜在基因(参见定位克隆)。

---

1) CpG 岛又称为 HTF 岛是由于它的片段中含有一簇 5-meC 敏感性限制性内切核酸酶 *Hpa* I 的作用位点。在正常的(即甲基化的)DNA 中,*Hpa* I 由于作用位点经常被切除或发生修饰而很少进行酶切。而在 CpG 岛中由于有大量的未发生修饰的作用位点因而 *Hpa* I 可以进行酶切从而产生 *Hpa* I tiny framents (HTF),所以 CpG 岛又称为 HTF 岛(参见限制性内切酶)。

在人类基因组中大约存在着约 50 000 个 CpG 岛。其中约有一半与管家基因 (housekeeping genes)相联系,另一半与细胞类型特异基因相联系。无论这些基因是否表达,这两类基因的中 CpG 岛都是组成型非甲基化的,失活 X－染色体上的基因则例外,它们的甲基化主要取决于亲本印迹 (在下文中讨论)。CpG 岛的位置和性质决定了未甲基化的启动子是转录所必需的,但并非是启动转录活性的充分条件。CpG 岛的人工甲基化将抑制转录活性也证明了这一点。启动子上的 CpG 岛可能由于与转录因子的组成型结合而避免了被 DNA 甲基化酶作用,这也可以解释非甲基化,转录因子的缔合以及异位甲基化效应。在细胞类型特异性基因中,可以认为某些转录调节复合物自发与 DNA 组成型结合 (产生了"岛"),转录的活化还要求其他的细胞类型特异因子。很多基因并不与 CpG 岛相联系。

哺乳动物细胞中的甲基化和基因调控　一系列的证据表明在 DNA 的甲基化和基因的转录,特别是在非**甲基化**或**甲基化不足** (hypomethylation)和转录活性之间存在着联系。例如上面讨论过的 CpG 岛对转录活性的影响,另外在转基因实验中被转化基因在进入细胞前如已甲基化则不能表达;甲基化阻断性药物如 5－氮杂胞嘧啶核苷对基因表达有诱导作用;以及细胞类型特异性基因甲基化模式以及在失活的 X－染色体上亲本印迹位点 CpG 岛的**超甲基化**(特定位点甲基化率上升)。

一般来说,甲基化的 DNA 能通过多种方式影响转录。首先,它能够直接干扰转录因子与其识别位点胞嘧啶的结合。其次,它能够与特异性识别甲基化 DNA 蛋白结合从而阻断或取代转录因子的作用,目前已有两种这种性质的**甲基化 CpG 结合蛋白**(methyl-CpG-binding protein)被分离出来。第三,它能够通过影响核小体的位置或与其他染色质蛋白相互作用而改变染色质的结构。

尽管已确知 DNA 的甲基化对基因表达有影响,但是除了亲本印迹和 X 染色体失活的例子(见下文),目前它在体内的调节机制还没有令人信服的证据。目前已有的大部分 DNA 的甲基化作为普遍调控方式的证据来源于细胞的体外培养系统,而这类 DNA 甲基化的模式与内源组织的甲基化模式常常并不相同。在体内可以观察到细胞类型特异的甲基化模式,有可能这种差异将导致基因活性的分化。甲基化通常被认为是基因表达差异的原因之一。例如调节蛋白与 DNA 的结合可以阻断维持甲基化酶作用于 DNA,从而引起基因的不同表达并维持这样的分化状态。实验发现,DNA 甲基化酶缺陷的纯合型小鼠胚胎在原肠胚形成后就迅速死亡,说明 DNA 甲基化酶对于体细胞的进一步发育是必需的。不过这种效应也可能是由于 X 染色体所有拷贝的失活而导致印迹基因产生轻度的下降调节(见下文)。

如果 DNA 的甲基化最终仅仅调控一小部分基因,那么在高等真核生物基因组中观察到的广泛的甲基化作用是什么? 一种理论认为它可以减少高等真核生物基因组中异常基因的表达,即它可以降低"遗传噪音"(genetic noise)。最近观察到的一个有趣的现象是在哺乳动物基因组中许多甲基化的胞嘧啶处于转座子中(参阅)。有两类*逆转录因子*(参阅),LINEs 和 SINEs(分别代表长和短的分散排列的逆转录因子)在哺乳动物 DNA 中含量特别丰富并且是哺乳动物基因组中中等分子量的 DNA 重复序列的重要组成 (见可移动遗传因子,基因组和作图)。DNA 的甲基化可以抑制植物和细菌的转座子,在脊椎动物中甲基化可能也有相同的作用。与之吻合,重组逆转录病毒转入小鼠的基因组后常常由

于超甲基化作用而保持沉默。有趣的是，由可移动遗传因子引起的染色体重排及突变的发生率在哺乳动物中非常低而在 DNA 不发生甲基化的果蝇中却非常高。

植物中的甲基化和基因的调控 当植物内源性基因的额外拷贝或待转化基因的多拷贝整合到植物细胞基因组后常常会导致内源和外源的基因发生外遗传沉默。这种外遗传沉默的机制可能发生在两种水平上：一种是在转录水平（即**同源性依赖转录基因沉默**，包括顺式或反式的 DNA 配对），另一种是在转录后水平（即**同源性依赖转录后基因沉默**或**共抑制**，包括反义 RNA 的作用［参见本条和**降解体**(degradosomes)］，但在这两种情况下 DNA 的序列都不发生变化。在许多例子下外遗传的沉默与序列中许多重复序列的超甲基化有关。

转录水平沉默所产生的遗传效应与**副突变**(paramutation)的自然过程十分相似。副突变是等位基因间的一种相互作用，即在杂合子中一条发生副突变的等位基因抑制了另一条能发生副突变的等位基因。在许多情况下这种副突变的结果是使这条能发生副突变的等位基因也发生副突变，并且这种副突变的效应可以通过减数分裂而遗传，但最终是可逆的。同样，副突变并不改变 DNA 的序列，而且这种效应似乎与基因的拷贝数和它在染色体上的位置有关。在许多但不是全部情况下副突变常常伴随着一定数量的 DNA 甲基化。

植物以及一些真菌中的重复 DNA 序列的超甲基化可能是一种阻止转座子传播的防御机制（与在脊椎动物中类似）。在共抑制和副突变中 DNA 重复序列所起的作用似乎也支持这种理论。与动物相似，植物在发育过程中经历全局性的甲基化状态改变，这可能会产生转位的高潮，很多转座子家族被迅速激活。例如一种阿拉伯果蝇(*Arabidopsis*)由于发生了 DNA 甲基化酶的缺陷突变，尽管能够存活，却存在着许多发育上的缺陷，可能对应于转座子家族的活化。目前 DNA 甲基化在植物中的普遍调控机制还不清楚，而 DNA 甲基化的缺陷可能反映了外遗传的降调节作用而不是突变发生的增加。

7.3 哺乳动物的 DNA 甲基化对外遗传基因的调控

甲基化的全局调控模式 在哺乳动物中，成体的特征的甲基化模式在原肠胚形成之前就已建立。在合子中甲基化的水平较高，而在卵裂期则发生广泛的去甲基化作用，随后从头甲基化发生仅在形成成体体细胞组织的外胚层。而胚外部分，包括胚外膜和原生殖细胞，在配子形成之前一直保持非甲基化，在形成后这时生殖细胞才发生从头合成甲基化。上面已讨论过，大部分甲基化和去甲基化作用都是组成型的，并且可能作为一种保护机制来抑制转座子的活性以及“遗传噪音”。但是在这中间也有两个例外，比如印迹基因和位于 X 染色体上的 *Xist* 基因座，它们在生殖细胞中以性别特异的方式甲基化，并且不经历发育早期发生的全局性去甲基化和从头甲基化。

亲本印迹 **印迹**(imprint)是记录外遗传信息的一种方式。**亲本印迹**是通过这一遗传机制可保留亲本来源的外遗传信息的染色体区域。在体细胞中，许多基因在母系和父系来源的染色体上同时表达，而印迹基因的表达却以亲本染色体特异的方式表达。通常，如果一条等位基因被抑制则它被称为被“印迹”（例如，在母系印迹基因座处，只有与它对应的

父系来源的等位基因能够表达)。但这只是一种较片面的定义,并不能够确切反映印迹机制的内在特征。在分子水平上,印迹就是 DNA 的甲基化,一些位点的 DNA 甲基化对应于基因的活化而另一些位点上则对应于基因的失活。在下文中我们将会谈到,许多印迹基因是成簇存在的,在顺式作用控制位点的单甲基化印迹就能直接导致一些基因的母系或父系的特异性表达。在遗传杂交中亲本印迹的结果如图 7.1 所示。尽管形成合子时父系和母系来源的染色体数量是相等的,但印迹基因座是功能性半合子式的,因此互交的结果并不一致。

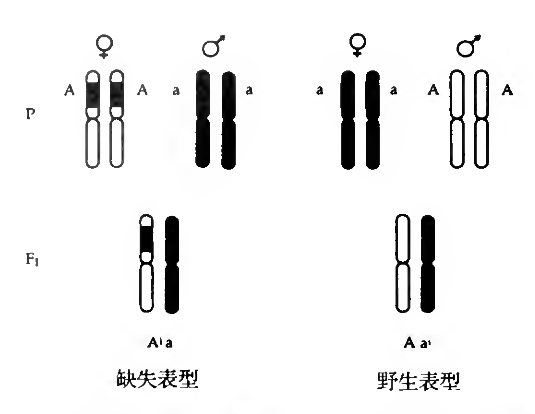


图 7.1 在互交中亲本印迹的效应。纯合的野生型个体(AA)与纯合的无效突变个体(aa)杂交产生杂合的 F_1 代(Aa)。母系等位基因携带印迹(深色条纹)并不表达,而父系等位基因表达。当母本为无效亲本时,F_1 代显示与父系等位基因 A 相关的显性表型。如果母系为野生型亲本,F_1 代显示与父系无效等位基因 a 相关的突变表型。

亲本印迹存在的第一个证据是从单性生殖发育(即由未受精的卵细胞进行发育)的失败中发现的。细胞核移植的实验表明父系和母系的基因组都是正常发育所必需的。单亲发育的小鼠胚胎,即由两只雄性(**孤雄生殖**,androgenotes)或两只雌性小鼠(**孤雌生殖**,gynogenotes)的细胞核组成的合子而来的小鼠胚胎在发育过程中常常表现异常并且在出生前就已死亡。其中孤雌生殖的胚胎可进行正常的早期发育,但它的胚外膜却并不发育。而孤雄生殖的小鼠尽管胚外膜发育正常,但它的胚胎却表现出生长停滞。这种反常的现象被称作 complete hytaditiform moles(partial moles 包括一种三倍体的胎儿,它由两个父系和一个母系的基因组共同组成)。使用含有同二体(isodisomic)二倍染色体片段的小鼠品系(染色体的两个拷贝都起源于同一个亲本)实验已经辨明了同基因印迹作用有关的染色体区域。例如,人类基因组上的同线(syntenic)(参阅)区域已被证实同一些典型的印迹异常疾病如 Prader-Willi & Angelman 综合征和 Beckwith-wiedemann 综合征有关。在动物中,印迹效应主要发生在哺乳动物中,在植物中也发现印迹效应,但由于植物对剂量有更强的耐受性,因而产生形态差异的表型而不是疾病。

亲本印迹和 DNA 的甲基化　目前在人类和鼠身上至少已辨明了 20 种印迹基因,而且其中许多编码与生长调节有关的产物。与鼠的印迹染色体区域一致,大多数人类的印迹基因集中在三个簇中。有证据表明在每个基因簇上都存在着特异的**印迹盒**(imprinting boxes)能顺式调节印迹基因的亲本特异性表达,这些位点上表现出亲本特异性的甲基化

模式，其建立与重新起始独立于哺乳动物发育的早期全局性的从头合成甲基化作用和去甲基化作用。在哺乳动物中，雄性和雌性配子中印迹盒发生的甲基化是不同的，从而能指导亲本特异性基因的表达。同时父系和母系染色体也由于重组的频率，染色质的结构以及复制时间的不同而表现出差异。

在人类染色体 11p15.5 位置上的印迹基因簇与胎儿的过度生长异常即 Beckwith-Wiedemann 综合征 (BWS)有关。这个基因簇中包含至少 7 个基因 (表 7.2)，在父系特异性的 IGF2 周围环绕着母系特异性的 *H19*，*KCNA9*，*CDKN1C* 和*NAP2*。*IGF2* 编码一种生长因子，它的降调表达在 BWS 综合征中起着主要的作用。在鼠的相关同源簇上也克隆到了三个印迹基因，其中 *Ins2* 是父系表达的印迹基因，它编码另一种生长因子且与人类基因簇上的 *INS* 基因有同源性，而人 *INS* 基因表达的模式目前还不清楚。小鼠中另外两个基因 *Mash* -2 和*Igf2r* 是母系表达的，其中 *Igf2r* 编码父系特异性的 *Igf2* 的受体 (人类 *IGF2R* 为双等位表达)。在鼠的某些组织中，*Igf2* 和*Ins2* 也表现为双等位表达，这表明基因的印迹作用是有种间特异性的，也就是说，这些基因是**被条件性印迹**(conditionally imprinted)。

表 7.2 人类 BWS 印迹基因簇和 X-失活中心(XIC)上的印迹基因

基　因	产　物	所处位置	表达情况
Beckwith-Wiedemann 综合征基因簇			
IGF2	生长因子		父系表达
INS	生长因子		未知——在鼠中为父系表达
H19	非翻译 RNA	11p15.5	母系表达
IGF2R	生长因子受体		双等位表达——在鼠中为母系表达
KCNA9	K^+通道		母系表达
CDKN1C	CDK 抑制剂		母系表达
X-失活中心			
XIST	RNA	Xq13	在原肠胚形成前为父系表达

注：在 BWS 印迹基因簇中生长调节因子占主要。*IGF2* 和 *INS* 编码生长因子。*IGF2R* 是生长因子的受体，*H19* 调节 *IGF2* 的转录，*CDNK1C* 被认为是细胞周期的调节子。鼠的 BWS 印迹基因簇上还包括一个母系表达的基因 *Mash* -2，它编码一个转录因子。

人类的 *H19* 和 *IGF2* 基因的互斥表达是调控 BWS 综合征的关键。许多 BWS 综合征的病例涉及到 *IGF2* 的双等位表达 (在正常情况下仅仅是父系染色体上的 *IGF2* 表达)，其中大部分反映了所谓的父系单亲二体染色体或父系三体染色体。然而，即使在正常的情况下两条染色体来源于不同的亲本，它的母系的 *H19* 基因座仍可能会发生缺失或甲基化，这是由于它包含了一个父系特异性的甲基化位点，因而其甲基化导致了 *H19* 的不表达。*H19* 突变体的上位作用和这种上位突变对 *IGF2* 转录的影响形成了**一种增强子竞争模式** (enhancer competition model)，在这个模式中 *H19* 和 *IGF2* 的转录被认为依赖于同一个增强子，当 *H19* 转录利用了这个增强子后就阻断了 *IGF2* 的表达。因此只有当 *H19* 发生甲基化作用阻止了转录后*IGF2* 基因才会进行表达。*H19* 的转录产物是非翻译的 RNA，除了转录能够抑制 *IGF2* 的转录外未显示其他的内在功能。

最近，包含完整的鼠 BWS 基因簇的酵母人工染色体(YACs)的转基因小鼠实验显示出反义的 RNA 在 *Igf2r* 的印迹表达中起着重要的作用(在正常情况下 *Igf2r* 仅在母系来源的染色体上表达)。在 *Igf2r* 的启动子上存在着一个父系特异的甲基化位点，在一个内含子中还存在着一个母系特异的甲基化位点，在人和鼠中这两个位点都存在，但在人类中这个基因是双等位表达的。*Igf2r* 基因内含子上的甲基化位点标明了反义链转录的启动子的位置，在未甲基化的父系来源的染色体上反义链能正常转录，从而抑制了父系的 *Igf2r* 的表达。当母系的反义链启动子发生甲基化，*Igf2r* 的转录去抑制。在上述这两个例子中亲本印迹的发生是通过不同的甲基化作用起始并由竞争机制来调节的。*Igf2r* 的这种竞争机制即反义链的转录所引起的逆向转录能对抗 *Igf2r* 的正常转录其本质目前还不清楚，它可能包被染色体并诱导类似与 *Xist* RNA (见下文)的异染色质的出现，或者可能存在一种类似于 *Igf2*/*H19* 调节元件的竞争机制。

印迹的作用：促进发育还是亲代间相互拮抗的结果？ 为解释哺乳动物中印迹的重要性和在进化过程中的作用提出了各种各样的理论。很多理论认为它在哺乳动物的发育过程中起着内在作用。实验发现虽然孤雌生殖小鼠不能正常发育，但是孤雌生殖细胞和正常细胞的嵌合体仍然能够存活。孤雌生殖细胞在异源嵌合的小鼠胚胎中的分布提示发生印迹分化的细胞在小鼠脑的发育中起着重要的作用。例如，孤雌生殖细胞在下丘脑中没有分布而在前脑中则大量存在。另一种**亲代拮抗模式** (parental conflict model)认为在多配偶的哺乳动物中印迹的出现是父系和母系基因组竞争的结果。在多配偶的哺乳动物中，父亲总是努力想使它的下一代获得最强的生存能力从而能战胜母亲与其他配偶产生的下一代，而母亲则总是想使更多的子代能活下来。因此，自然选择导致来自父系的等位基因能促进个体的生长，而来自母系的等位基因则抑制个体的生长。在鼠的 BWS 基因簇中，父系的 *Igf2* 和 *Ins* 编码生长因子，而母系的 *H19* 则是 *Igf2* 转录的抑制子，同时母系的 *Igf2r* 作为 *Igf2* 的清除受体能够降低血清中生长因子的水平。在 *H19* 或 *Igf2r* 被敲除的小鼠中由于 *Igf2* mRNA 和游离蛋白水平的增高促进了循环的 *Igf2* 数量的增加，因此小鼠出生时的体重也呈增长的趋势(相反的如果来自父系染色体的 *Igf2* 等位基因被敲除的话，其生长速度只有正常小鼠的 60%)。有趣的是，当一种多配偶的老鼠和另一种单配偶的老鼠交配后，它们 F_1 子代出生时的体重也表现出亲本印迹的现象。当多配偶的公鼠与单配偶的母鼠交配后它们产生的子代要比单配偶的公鼠与多配偶的母鼠产生的子代要大。这种现象符合亲代拮抗的模型。因为单配偶的公鼠生育更强壮的子代的压力较小而单配偶母鼠促进更多子代存活的压力也较小而当这两种不同的鼠种间进行交配时，由于多配偶鼠的竞争策略使 F1 子代体重呈上升的趋势。综上所述，这种模型只是预言了而非反映了哺乳类发育的内在性质。比如 Igf2 和 Igf2r，或者是 Igf2 和 H19 被敲除的小鼠，它们的表型都和正常小鼠一致。

X 染色体失活 在雌性哺乳动物的体细胞中，X 染色体的剂量补偿(参阅)可以通过 X 染色体的失活来实现。一条 X 染色体缩聚成异染色质而失去转录活性。失活的 X 染色体(**Xi**)在整个细胞周期中一直保持缩聚的状态，可以被染成深色，称作**巴尔小体**(Barr body)，而另一条活化 X 染色体(**Xa**)则保持正常。

在雌性的配子中，起初两条 X 染色体都具有活性。X 染色体的失活首先发生在胚外

系的囊胚泡期的晚期以及在更晚一些时候发生在形成了胚胎体细胞的外胚层系统中。在具有胎盘的哺乳动物胚胎组织中，失活的 X 染色体的选择是随机的。一旦被选中这个过程就不可逆转：具有活性和失活的 X 染色体分别克隆扩增，所以在一个杂合子个体中可见的 X 连锁的标记就能够显示被标记表型的花斑分布（例如玳瑁猫的毛皮）。在许多哺乳动物的胚外组织中（但不包括人类），父源的 X 染色体更容易失活（亲本印迹的一个例子，见上文），而在有袋类动物的胚胎组织和胚外组织中却倾向于父源的 X 染色体失活。在这两个例子中，亲本印迹不是 X 染色体失活所必需的，因为在包含有两条母源 X 染色体的胚胎中也能发生正常的 X 染色体失活现象。

X 染色体失活的分子机制　　X 染色体的失活可以分为四个时期：计数期，起始期，增殖期（传播期）和维持期。一个 X 染色体的顺式作用位点，**X 失活中心**（*Xic*），似乎与这四个时期都有关，通过对 X 染色体的缺失和转位所引起的失活破坏的实验而定位了这个位点。如果含有 *Xic* 的区域被缺失，突变后的 X 染色体仍然具有活性；而如果 X 染色体发生了转位，则包含 *Xic* 的片段会失活（转位到常染色体构成常染色体的一部分），而其余的片段仍具有活性。

通过 *Xic* 的精确定位发现了一个单基因座，*Xist*，它仅仅在 Xi 上有特异性的表达（即 Xi 特异转录体）。*Xist* 编码一种很大的非翻译的 RNA，它在细胞核内与 Xi 相互结合并被认为通过包被染色体而诱导染色体的异染色质化。几个设计巧妙的实验提供了清晰的证据说明 *Xist* RNA 在 X 染色体失活过程中所起的重要作用。例如，当转染的 *Xist* DNA 整合到雄性 ES 细胞（参阅）的常染色体上后导致了重组的常染色体或内源性 X 染色体的随机失活（也就是说含有 *Xist* 的常染色体被细胞认作第二条 X 染色体）。在鼠中，另一个基因座，*Xce*（X-controlling element）与 *Xist* 相连并且能够影响 X 染色体失活的随机性质，通常情况下，含有相对较弱的 *Xce* 等位基因的 X 染色体更容易失活。

计数期和起始期　　X 染色体失活的总体调节反映了 X 染色体和常染色体之间的一种平衡关系。多体性 X 染色体（参阅）由于额外 X 染色体的失活产生了多个巴尔小体而耐受，而多倍性（参阅）则导致了更多的保持活性的 X 染色体。计数期机制的本质目前还不清楚，而转基因小鼠的研究对 *Xce* 所起的作用提供了互相矛盾的证据。目前还不清楚当多个顺式的 *Xce* 转入后是否是单个计数的，在具有二倍 *Xce* 的人类男性中并不发生 X 染色体的失活。

在配子发生过程中，*Xist* 启动子发生性别特异性的甲基化作用，从而在合子中产生了母系印迹（见上文的亲本印迹）。这导致了父源 *Xist* 等位基因的优先表达，而引起了父源 X 染色体的失活倾向。在鼠的胚外系统中这种甲基化的模式能够维持（父源的 X 染色体保持失活状态），而在鼠的外胚层系统中这种甲基化的模式由于一种全局性的去甲基化作用而重新起始，而 *Xist* 启动子的从头合成甲基化作用则随机发生活性与失活染色体的选择最终要依赖从头合成甲基化酶，但是目前仍然不清楚这种随机的选择是如何进行的。在人类，X 染色体的随机失活在胚外系和外胚层系中都有发生，在一些胚外细胞中两条 X 染色体都保持活性。

有人认为细胞内 X 染色体的失活倾向性是对染色体突变的回应。因此杂合子个体中 X 染色体较大区域的缺失将导致缺失染色体的失活，从而能够避免以正常染色体为代

价去维持这样的缺陷带来的致命后果。事实上，细胞在染色体的选择上并无偏见，观察到的细胞的均一性反映了细胞的这种选择。细胞存活能力的差别是由于在发育的早期细胞含有不同的 Xi 和 Xa 的结果。当细胞含有发生突变的 Xa 时不易存活，而只有含有正常的 Xa 的细胞才能扩增形成完整的胚胎。

X 染色体失活的增殖期和维持期　Xi 的失活反映了染色质结构和 DNA 甲基化模式的改变。当乙酰化的组蛋白 H4(见第 3 章染色质)缺失时，染色质是高度缩聚的，而且许多 X 连锁基因上游的 CpG 岛是超甲基化的。引起这些变化的信号目前还不知道，而 X 染色体沉默中涉及的染色质蛋白也还未被识别(参考花斑位置效应，SIR 蛋白质)。*Xist* RNA 自身可能在对染色质的控制中起着重要作用，但它还不足以引起 X 染色体的完全失活：在 Xi－常染色体转位中，失活可能通过转位的转折点而扩散进入常染色体但决不作用于常染色体的末端，尽管 *Xist* RNA 包被了整个染色体。这说明在常染色体上可能存在着一种边界元件阻断了失活的扩散，或是阻断了 X 染色体上顺式失活部位，**激发元件**(booster element)。

在正常的 X 染色体上，失活作用从 *Xic* 传播需要好几天并且失活在 CpG 岛超甲基化前就已发生。失活作用的传播可能代表了一种协同效应即特异性的蛋白结合到已存在的失活染色质上，这与异染色质在常染色体转位中的传播相似(参考花斑位置效应)。在人类，一些 X 连锁的基因能够"逃脱"失活（例如编码类固醇硫酸酯酶的基因)。这样的基因常常成簇存在，说明在 Xi 中存在着一个局部的活性区域。由于在鼠中对应的基因常常是失活的，因此这一现象无法外推到其他的哺乳动物。看来 DNA 的甲基化在 X 染色体的失活中发挥着双重的作用：首先可以引起 *Xist* 的表达，其次可以维持转录的抑制作用，也就是说，维持甲基化酶的活力使失活的 X 染色体通过随后数次的 DNA 复制进行克隆扩增，并维持其失活状态。

(王家楠 译)

第8章
基　因

基本概念和定义

- **基因**(gene) 是遗传信息的物理和功能单位,它具有被表达潜势,即可被用作模板产生一种或多种 RNA 或蛋白质的基因产物。基因组 (参阅) 指细胞内所有遗传信息,这种遗传信息以核苷酸序列的形式存贮。所有的遗传信息都可以作为染色体的一部分被存贮并世代相传,但只有基因被表达,其他基因座的功能仅处于 DNA 水平上,其中包括调控元件,复制起点,着丝点和端粒。
- **顺反子**(cistron) 是由互补试验定义的遗传功能的单位,例如,细菌基因就等同于顺反子——它是对应于一个开放阅读框的遗传功能单位。但在真核细胞中,基因与顺反子的关系较为复杂,这是因为真核基因可被选择性使用,或是在转录后发生改变,还可能同其他基因一起产生一种基因产物。

8.1　基因的概念

基因概念的发展　　基因一词最早在 1909 年为 Wilhelm Johansen 采用来描述传递和表达特定的生物性状的可遗传因子,并未涉及任何特殊的遗传理论。可见,基因最初的含义中不包含特殊的物质基础,纯粹作为一个抽象的名词使用,如同在传递遗传学中使用的情况 (参见孟德尔遗传)。

20 世纪前半叶,几个独立的工作导出了基因物理和功能基础的更为精确的概念。1902 年,Archibald Garrod 提出,代谢紊乱中的尿黑酸症是由于某一种特殊的酶的失活引起的,并且以常染色体隐性遗传的方式传代。Garrod 将这种现象称为**先天性代谢缺陷**(inborn error of metabolism), 由于 Garrod 并不熟悉孟德尔遗传,这个发现的重要性直到 30 年后 George Beadle 和 Edward Tatum 发现了用 X 射线诱导的真菌突变导致特定的生化缺陷才被认识,这些发现促进了基因功能的**一基因一酶**(one gene one enzyme)模型的出现。1911 年,Thomas Hunt Morgan 指出基因定位于染色体上并且在物理上相连 (参见遗传的染色体学说)。1944 年,Oswald Avery 和其同事通过消除实验证明,DNA 就是遗传物质,这样便勾勒出基因概念的轮廓——即染色体上的一段长度的 DNA,它可以编码一个酶的信息。

基因这个概念必须进一步扩展才能涵盖现在所了解的关于基因功能的知识:首先,不是所有的基因都编码酶(一些基因编码具有不同功能的肽链,有些基因编码功能性 RNA 分子,

如 tRNA 和 rRNA);其次,在病毒中,基因可能是 RNA 而不是 DNA。更进一步地,基因中的信息可被选择性使用产生一种以上的产物。孟德尔遗传中关于单个基因控制单个生物性状的概念也有所扩展,现在已认识到一个基因可能同时影响多个性状(**多效性**,pleiotropy),而多个基因可以相互合作控制同一性状(参见数量遗传)。随着分子生物学的出现,现在可以更精确地从基因结构和功能来定义基因概念,但是由于真核和细菌在基因表达和基因组的组织上的策略有显著差异,因而基因的定义有所不同(参见朊病毒)。

8.2 遗传结构和遗传功能的单位

遗传结构的不可分割单位 在基因位于染色体上并且能够用重组的方法作图定位(参见遗传作图)的概念建立后,单个基因逐渐被认为是遗传信息结构和功能的基本和不可分割的单位,从结构和功能来看,它们以线性的形式相互连接(**串珠理论**,the beads on a string theory)。然而这种观点又受到了噬菌体重组实验结果的挑战,这些实验结果表明,基因可被分为更小的单位。于是,Seymour Benzer 引入了**突变子**(muton),**重组子**(recon)和**顺反子**(cistron)这些术语分别定义突变,重组和功能的不可分割单位。在噬菌体感染中,通过同一基因的不同突变的杂交实验可以恢复产生野生型的噬菌体。如果突变位于同一基因中不同亚元件中,那么,这只可能是**基因内重组**(intragenic recombination)的结果。这说明基因可被分为更小的单位,这些单位可发生重组和突变。这样,重组子和突变子等价于单个核苷酸对。

基因作为遗传功能的不可分割单位 Benzer 的顺反子概念意为基因功能的不可分割的单位。顺反子的确定可通过互补分析实验(complementation analysis)确定。在实验中,测试基因(或更特指其产物)在同一细胞中补偿同源基因突变的能力。成功的互补可恢复野生型表型。

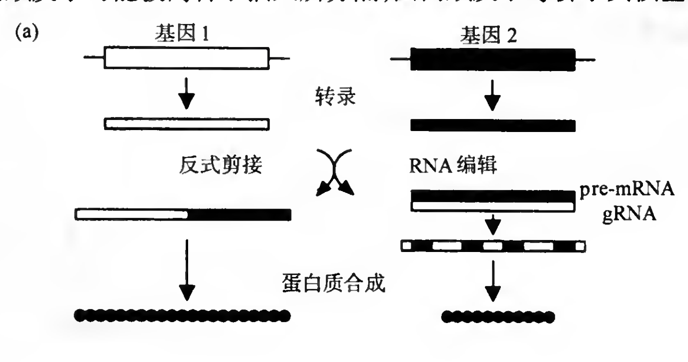
互补实验的基础是**顺反测验**(*cis-trans* test)(顺反子一词来源于这个实验;文框 8.1),成对的独立的突变被认为存在顺和反两种构型。顺式测验是对照组,因为如果两个突变均在同一个基因组中,那么另一个基因组的两个基因座均为野生型,其产物为正常的基因产物,这样细胞表现出野生表型。反式测验是互补实验,可以确定功能单位的边界。如果两个突变在同一个基因中,那么它们以反式构型出现在细胞中时,每一基因组都携带有这一基因的突变体拷贝,因而在细胞中不能产生具有功能的产物——即不出现互补。如果突变位于不同基因中,当它们以反式构型出现时,那么每个基因组均可补偿另一个基因组缺少的正常产物。细胞具有所有基因产物,表现为野生型——这就是**正互补**(positive complementation)。

在细胞和酵母中的互补分析建立了一基因一顺反子的概念,即基因可被定义为功能单位。这种实验在指定途径中确定参与基因的数量时也十分有用,而且还可进一步通过突变体之间的互养实验来确定基因之间的作用顺序。此外,互补实验还可以用于探究基因功能或克隆已知功能基因(参见标记获救、功能克隆、表达克隆)。

8.3 原核和真核细胞中基因-顺反子的相互关系

在简单基因组中基因与顺反子等价 在原核和低等真核细胞中,基因同其产物之间的关系一般比较简单。在许多情况中,通常是一基因一相应产物,而且基因往往与产物共线性。因而,在这些生物中,基因和顺反子等价:基因是遗传功能单位;也是可表达的遗传信息的单位。在细菌中,基因是编码区的同义词(开放阅读框),而在真核细胞中,基因是转录单位的同义词,这是由于细菌基因常常组合成一个操纵子(参阅),这样几种产物均由一条**多顺反子 mRNA** (polycistronic mRNA)翻译而成,相反,在真核细胞中,大多数基因以**单顺反子 mRNA** (monocistronic mRNA)的形式转录(参见 rRNA 基因、反式剪接、内部核糖体进入位点)。

复杂基因组中基因与顺反子不等价 在高等真核细胞的基因组中,基因和产物之间的关系较为复杂(图 8.1)。大多数高等真核细胞基因包含有内含子,它们是一些不出现在最终产物中因而不是功能组成部分的 DNA 插入序列(见第 12 章基因组与作图)。真核基因代表整个转录单位,而顺反子可能被内含子插入所分隔,因而顺反子等价于真核基因的外显子。



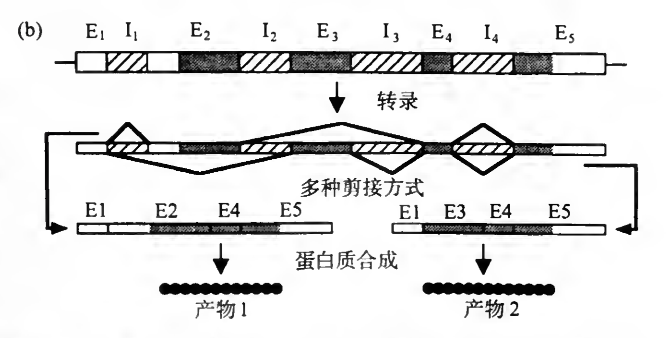


图 8.1 在真核细胞中基因与顺反子不等价的例子(图中较宽的框表示 DNA,窄框表示 RNA,链表示蛋白质)。(a) 在反式剪接与 RNA 编辑的例子中,单个多肽链的合成需要多个基因的表达,每个基因都是同一功能单位的部分并构成单个顺反子。空白的框代表来自某一基因的信息而填充的框代表来自另一基因的信息。注意在所有已知的反式剪接的例子中,5′剪接转录物不被翻译,虽然在理论上还不能解释它为何不产生蛋白产物。(b) 一个基因通过多种剪接方式或其他选择性信息利用方式(见正文)产生多种产物。基因中包含相互覆盖的顺反子。内含子用有斜纹的框表示,它们在 RNA 加工过程中被剪切。外显子也用框表示,如果不被翻译用空白框表示,填充框表示编码区。注意内含子可以插入到编码或非编码的外显子中,而外显子可以包含翻译或非翻译信息(即外显子 2 和 5)。

在真核基因中，基因与产物相互关系的复杂性还来源于某些遗传信息被选择性地利用以产生多种的产物。这种过程可通过选择性剪接（参阅）来完成，这种选择性反映了在RNA加工过程中或在启动子选择，及转录过程中多聚腺苷酸位点的使用等水平的调节（见第27章RNA加工）。这些结构上相联系的基因产物往往具有不同的功能，因而基因还可能包含一系列的相互重叠的顺反子。

同上述情况相反，有时一种产物需要两个基因共同产生，如反式剪接，即两个分别编码的mRNA被剪接在一起翻译，另一个例子是在锥体虫中的RNA编辑（参阅），mRNA和gRNA都是产生蛋白合成的成熟模板所必需的。这种情况中每个基因对于产生共同产物都是必需的；它们是同一顺反子的一部分。

还有一些例子中几种不同蛋白都来源于同一个开放阅读框：翻译首先产生一个**多蛋白**（polyprotein），随后在被剪切成具有不同功能的产物。一些RNA病毒采用这种策略以适应真核细胞中单顺反子的机制。这种情况也发生在一些内源基因中，例如在哺乳动物的脑中，前强啡肽原基因可以产生七种有着不同功能的多肽。在这种情况中，编码每种肽段的开放阅读框被认为是一个顺反子。

重叠和嵌套基因　大多数基因是分立不重叠的单位，即它们不和其他基因使用共同的遗传信息。**重叠基因**（overlapping genes）指调控具有独立性但部分使用共同基因序列的基因。大体上，基因重叠发生在两种水平。在细菌系统或其他的空间限制必需的情况中（如在RNA病毒基因组中和动物的线粒体DNA中），基因可以在阅读框水平上重叠，这样同样的遗传信息产生两种或多种互不相关的蛋白质。在同一开放阅读框可将负链为模板转录，或从不同的方向翻译，可以是不同基因使用不同的读框，也就是说基因间除了共同使用部分空间以外，并无相似之处。其中一个例子便是轻小病毒的裂解蛋白基因（包括噬菌体MS2），与复制酶和衣壳蛋白基因重叠，但是它们从不同的方向和以不同的开放阅读框翻译。在一些种属中，裂解蛋白基因完全包含在复制酶基因中。在真核细胞中，基因在转录单位的水平中相互重叠，但是外显子保持分立。由于在一个基因中的外显子DNA被作为其重叠基因内含子的一部分，因而两个基因的蛋白产物中并不出现相同的信息（例如，人的*TCRA*和*TCRD* T细胞的受体基因在外显子水平上重叠）。偶然还会出现一个完整的基因包含在另一较大基因的内含子中：与内含子代谢有关的编码蛋白的开放阅读框往往位于自剪接内含子中（见第27章RNA加工）。例如有3个小基因就隐含在较大的人类基因NF-1的第26号内含子中。重叠基因可能反映了一类调控机制。在质粒中，基因编码的反义RNA往往同其调控的基因相重叠（参见反向转录）。

嵌套基因（nested gene）指那些通过调节蛋白合成终点而产生两种或更多种嵌套蛋白产物的基因。这可能通过终止子的渗漏通读（例如Qβ病毒的衣壳蛋白基因），或者是翻译时发生移码（如在大肠杆菌*danX*基因和F质粒的*traX*基因）。真核细胞的RNA病毒也采用类似策略如反转录病毒，并且真核基因也可通过选择性剪接或采用选择性多聚腺苷酸位点产生嵌套产物。

8.4　基因结构和构造

基因的结构组成　基因可划分为具有特定功能的分立区域（表8.1）。

表 8.1 定义基因功能性成分的术语

术语	定义
等位基因	基因的一个序列变异体（或者称为遗传标记，如 RFLP，VNTR 序列）。
顺反子	基因功能的一个单位，可以编码特定产物的一段 DNA。
编码区，开放阅读框（ORF）	可以翻译成蛋白质的 DNA 区域，在细菌中，即为一个基因。在真核细胞中，编码区可以被内含子隔断。
分段基因	具有处于不同基因座上的外显子区的断裂基因，这些外显子必须分别转录并通过反式剪接相连接。实际上每一基因座应被认为是一分立的基因，这属于术语的误用。
基因	在细菌中，是指编码一个分立的蛋白质或 RNA 分子的遗传功能单位。在真核细胞中，是指编码一个或多个产物的，或对某一个产物产生有贡献的一个转录单位。
基因座	染色体上一个基因的位置，包括两侧的调控元件。基因座一词的本义是指任何标记物的位置——包括基因，调控元件，复制起始区，细胞遗传学中的标记等等。
操纵子	包含几个基因（可作为一个多顺反子的转录物被转录）及其共同调控元件的细菌基因座。
假基因	类似基因的一段无功能的序列（见第 15 章突变与选择）。
被分隔基因	包含内含子的基因。
转录间隔区	RNA 基因或 RNA 基因操纵子中不出现在成熟 RNA 分子中的部分。
转录单位，转录区域	可以转录为 RNA 的一段 DNA 区域。在真核细胞中即为一个基因。在细菌中可能包含多个基因。
非翻译区（UTR），非编码区（NCR）	转录单位中不能翻译成蛋白的部分。在编码区或操纵子两侧的 UTRs 为 5′和 3′UTRs（或称为前导和尾随序列）。

对于任何给定的基因座，被转录的 DNA 称为转录单位（transcription unit）。在原核细胞中，一个转录单位可能包括多个基因组成一个操纵子，但在真核细胞中，转录单位几乎总是等价于单个基因（由 RNA 聚合酶 I 转录的 rRNA 基因多顺反子，RNA 病毒基因和细胞器基因组等例外；参见内部核糖体进入位点、反式剪接）。

对于编码蛋白质的基因，翻译成多肽序列后的信息和未翻译的信息之间可能存在差别。在细菌中，被翻译的区域[**开放阅读框**（open reading frame）、**编码区**（coding region）]同基因等价，并且，基因间通常被短的**内部非编码区**（internal noncoding regions）分隔。操纵子的末端基因的侧翼存在有 **5′非翻译区**（5′untranslated region，UTR）或称为**前导序列**（leader sequence）和 **3′UTR** 也称为**尾随序列**（trailer sequence）。这些序列往往具有调控功能；5′UTR 控制核糖体的结合还可能促进**衰减子控制**（attenuator control）；而 3′UTR 在 mRNA 的稳定性中起重要作用。在真核细胞中，编码区的两侧也存在具有调控功能的 UTRs，两侧的 UTRs 和开放阅读框都被非编码序列即内含子插入，内含子在 RNA 从核仁运输出来时被剪切掉，也就是说它们不出现在成熟的转录物中。

在真核和原核细胞中，RNA 基因可被单独或作为操纵子的一部分转录。基因中与蛋白编码区类同部分即最终形成成熟 RNA 的部分。一些 RNA 作为成熟转录物被转录，而

另一些需经过剪切,加工和内含子的剪接等过程。渐次丢失的所有序列都被称为**转录间隔序列**(transcribed spacer sequence)。细菌和真核细胞基因的典型结构组成见图 8.2。

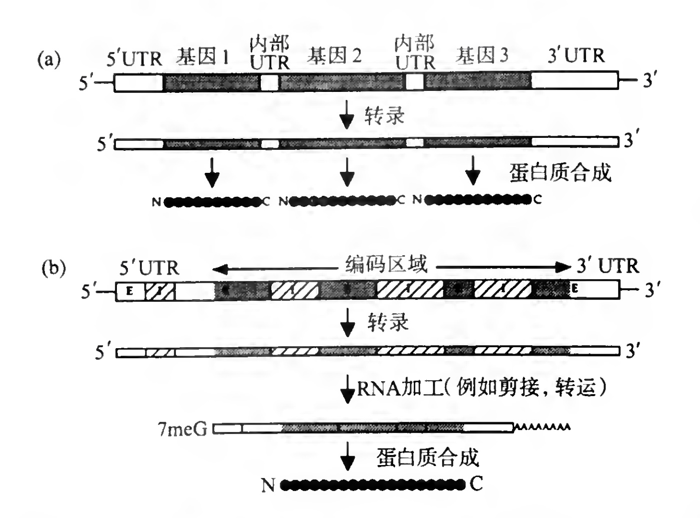


图 8.2 原核细胞与真核细胞编码蛋白质基因的典型结构(图中较宽的框表示 DNA,窄框表示 RNA,链表示蛋白质)。(a) 在细菌中,同一转录单位包含多个基因。它们作为可分为不同编码区(基因)的同一 mRNA 被转录,编码区在图中用灰色表示,非编码区用白色表示。每一基因编码一独立的多肽链。(b) 在真核生物中,转录单位由单个基因组成。基因包含中心的编码区域,用灰色表示,两侧的非编码区域用白色表示。基因往往被一个或多个内含子(用斜纹表示)插入,内含子在转录完成后被剪切。在图中的例子中,成熟的转录物编码单个多肽链,但不同的剪接方式可能可产生一系列不同的产物。

基因命名法　基因的命名一般根据种属习惯(表 8.2)。一般用斜体表示基因的名称,等位基因及其基因型,或在必要时表示基因转录而成的 mRNA,而蛋白产物和表型用正体来表示。但是在研究不同生物的同一遗传机制时,往往会产生一些混淆,如在研究酿酒酵母和粟米酵母的细胞周期有关基因的命名中(见第 2 章细胞周期)。此外,许多基因在不同实验中从相同组织被分离出好几次而具有不同命名:重要的果蝇的发育基因 *torpedo* 便是其中一例——它在筛选不同表型的过程中三次被鉴定并被命名三种不同名称。果蝇提供了关于遗传命名的最为丰富的例子,特别是在发育生物学中(见第 6 章发育,分子机制),这种趋势也扩展至脊椎动物中。

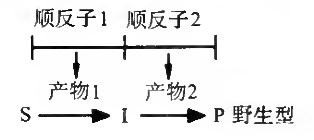
在许多种属中,基因由包括几个字母和数字的符号来表示,一些种属命名惯例 (如果蝇、大肠杆菌)认为使用小写字母表示隐性突变,而用第一个字母大写来表示显性突变。在其他一些种属包括人的基因命名中,基因全由大写字母表示。现在,通过大规模的测序方法,更多的基因不断被鉴定,因而十分需要一个统一的命名方法。

表 8.2 传统基因命名法提要

物种	惯例
大肠杆菌和其他细菌	三个小写字母表示一个操纵子,接着的大写字母表示不同基因座。例如:*lac* 操纵子;基因座:*lacZ*, *lacY*, *lacA*。蛋白质:LacZ, LacY, LacA。另外还采用特殊惯例命名 *B. subilis* 的孢子发生基因(见文框 6.2)。这些基因以 *spo* 后加上表示孢子发生的形态阶段的罗马数字表示,再用大写字母表示操纵子,而后为基因座,例如,*spoIIGA* 就表示在第二阶段表达的操纵子 G 的第一个基因座。
酵母	三个字母表明基因功能,而后的数字表示不同的基因座。啤酒酵母基因 *GAL4*, *CDC28*;蛋白质:GAL4, CDC28。非洲粟酒酵母基因 *gal4*, *cdc2*;蛋白质:Gal4, Cdc2。
线虫	用三个小写字母表示突变表型,如存在不只一个基因座,用连字符后接数字表示,例如,基因 *unc-86*, *ced-9*;蛋白 UNC-86;CED-9。
果蝇	来自突变表型的描述可以用 1~4 个字母代表。例如,基因 *white*(*w*), *tailless* (*tll*), *hedgehog* (*hh*);而蛋白为 White, Tailless, Hedgehog。
植物	虽然没有适用于所有植物的惯用法,但大多数用 1~3 个小写字母表示。*Arabidopsis* 基因用果蝇的方法命名但使用大写字母,例如,基因 AGAMOUS(AG),蛋白 AGAMOUS。
脊椎动物	一般以描述基因功能的 1~4 个小写字母和数字表示其基因功能。例如,基因 *sey*, *myc*,蛋白 Sey, Myc。
人类	方法如脊椎动物但需大写。例如基因 *MYC*、*ENO1*,蛋白 MYC、ENO1。

文框 8.1 顺反测验:

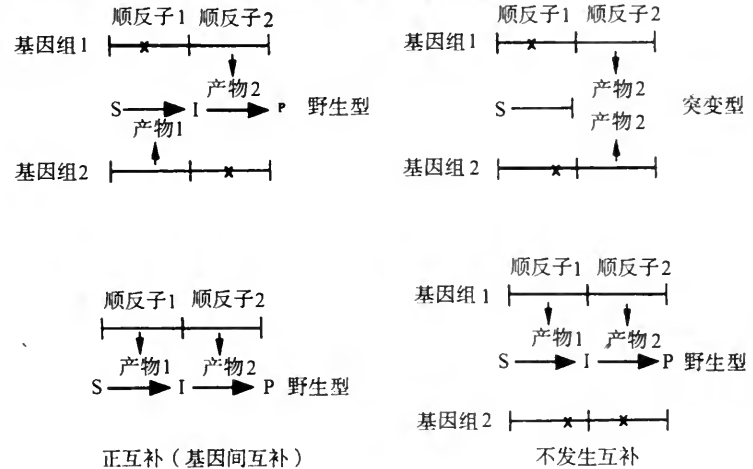
实验原理 顺反测验可应用于在同一个细胞中的成对突变,来验证相互互补是否会恢复野生表型(是否每个基因组可补偿另一个基因组的缺陷)。现举例说明,如果两种基因产物都是将底物(S)经中间物(I)转化成色素(P)所必需的,而这种色素是野生型的表型特征之一。当不能产生这种色素时将导致突变表型,它同野生表型从颜色上即可区分。正常路径用下图表示。



反式测验 实验中的反式部分是互补测验。两个突变被认为处于反式构型中,**正互补**(Positive complementation),也即野生表型的恢复通常只发生在两个突变处于不同的顺反子中的情况,也就是说,二倍体细胞至少携带每种基因野生型的一个拷贝。这被称为**顺反子间互补(又称基因间互补)**(intercistronic (or intergenic) complementation)。如果突变位于同一个顺反子中,互补通常

不会发生，因为两个基因组都没有野生型的基因拷贝，因而每个基因组都不能产生有功能的产物。

等位互补　在一些偶然情况中，位于同一个顺反子中的不同突变相互间也可以发生互补，这便是已知的**等位互补，顺反子内互补或称基因内互补**（Allelic, intracistronic or intragenic complementation）。这种情况发生于基因产物需以多聚物的形式作用时，一个突变基因产物可以补偿多聚复合物中另一个产物的缺陷。由于突变的异多聚物比野生型的同多聚物活性低，等位互补通常可以同正常基因间互补区分开。一个实际例子便是 **α-互补**（α-complementation）：许多克隆载体都只携带有大肠杆菌 *lacZ* 基因的 5′部分，它不能产生有功能的酶。这些载体可在缺失 5′ *lacZ* 的宿主细胞中使用，在细胞中，两种 *lacZ* 的缺失等位基因都没有功能，但是两个蛋白部分可以缔合形成有功能的酶。α-互补多应用于基因克隆实验中的重组筛选中(参见蓝白斑筛选)。



正互补（基因间互补）　　不发生互补

顺式测验　测验的顺式部分仅仅是一个对照，因为当两个突变位于同一个基因组中，那么第二个基因组是野生型，因而细胞显示出野生表型。当这种现象不出现时，这意味一个基因组产生的突变基因产物对于其他基因组产生的野生产物起显性作用，这称为**负互补**（negative complementation），与等位互补类似，这种现象在基因产物是多聚物时常常发生。在这种情况下，失活的突变产物同野生产物作用产生没有活力的异多聚物，例如，它可将野生型多肽链束缚于一个失活的复合物中。这种突变又被称为**反式显性**（*trans*-dominant）或**显性失活**（domimant negative）。

顺式显性　互补仅发生于基因之间，这是因为它们产生可扩散的基因产物。在同一个细胞中，即使提供了野生型的等位基因，一些突变也不能发生互补，这又称为**顺式显性**（cis-dominant）。这表明存在顺式作用元件的突变而非基因本身的突变所造成。

极性突变　细菌互补分析中的复杂性还体现在**极性突变**（polar mutation）的效应中。由于很多细菌基因排列于操纵子中，而且以多顺反子 mRNAs 形式被转录，核糖体的结合依赖于上游基因翻译的完成，造成一个基因的突变可以干扰下游基因的表达。

互补，回复突变和重组　上述三个过程均可恢复野生型。回复突变是突变的逆过程，从功能丧失的突变恢复成野生型的等位基因，即它包括基因型的改变。在两个突变的等位基因间的重组也可以产生野生型的基因组，如同双突变，它也涉及了基因型的改变。而互补不包括基因型的改变，它反映了基因产物间的相互作用。在互补分析中，回复突变和重组会产生似是而非的结果，但是

回复突变和重组都可以根据它们在分离后保持野生表型而被区分。由于重组造成的复杂性可用使用重组缺陷型的株型的方法避免。

互补作图 相互间不发生互补的突变称为一个**互补群**(complementation group),它们对应于同一顺反子。进行许多突变间成对互补分析可构建一个**互补图**(complementation map),它以分立不重叠的线段表示互补组构成。这个方法可以确定特定生化途径中包含基因数量,而且还可通过互养(将一个突变体暴露于另一的突变体的产物中)确立基因的作用顺序。在严格意义上,互补图不能称为基因图谱,因为它没有对基因间的物理联系加以说明。

互补作图的一个简单例子如下所示。五株均不能合成一个特殊代谢物的酵母进行互补实验,其实验结果如左下图所示:+ 代表正互补(即野生型的恢复),- 代表没有互补。图中表明 1、2 形成一个互补群,3、4、5 形成另一个互补群即存在第二个基因。在第二基因中,4、5 可以互补,这是等位互补,因而第二个基因的产物很可能以多聚物形式作用。

	1	2	3	4	5
1	-	-	+	+	+
2	-	-	+	+	+
3	+	+	-	-	-
4	+	+	-	-	+
5	+	+	-	+	-

1,2　　3

4　　5

(袁　翀　译)

第9章
基因表达和调节

基本概念和定义

- 基因(参阅)不同于其他形式的遗传信息,因为它需要被表达。**基因表达**涉及从基因中得到信息,并利用它产生**基因产物**,产物可以是 RNA 或蛋白质。
- 以最简单的形式,基因表达可以总结为分子生物学的**中心法则**(central dogma):遗传信息单向从 DNA 通过 RNA 再到蛋白质(表 9.1)。这对大多数细胞内基因都适用,但在病毒中信息在核酸之间的传递较为复杂。无论如何信息不能从 DNA 直接到蛋白质或蛋白质返回到核酸。
- 基因产物最终释放之前,基因表达通过不同环节进行。主要的环节发生在转录和蛋白合成,但还存在很多对遗传信息进行过滤和修饰的微调过程。原则上所有阶段都能被调节。基因表达各阶段的主要调节方式在真核生物和原核生物中有所不同的。
- 基因调节用于控制细胞产生基因产物的量。调节涉及顺式作用元件和反式作用因子;可以是正调节或负调节;诱导或抑制;广泛的或专一性的。这些术语将在下文中讨论。

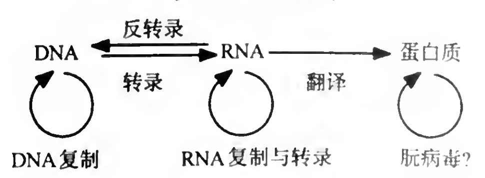
表 9.1 原来的中心法则和考虑到遗传信息流向的多种途径而作的修正

中心法则

DNA —转录→ RNA —翻译→ 蛋白质

中心法则的延伸	举例
DNA→RNA(没有蛋白质)	RNA 基因
RNA→蛋白质(没有 DNA)	RNA 基因组
RNA→DNA(反转录)	类逆转录病毒复制
RNA→RNA(RNA 复制,RNA 转录)	RNA 病毒复制和转录

修正后的中心法则



9.1 基因表达

基因表达的多个阶段 基因表达的阶段(表 9.2)即信息从基因到产物转化途径中的可以被调控的独立阶段。转录常被认为是基因表达的起始阶段,并是基因调节的主要阶段,但在转录成为可能之前,基因必须进入适于转录的形式。真核生物中准备阶段特别重要,在这个阶段中,DNA 甲基化和染色质中 DNA 的包装对其转录能力有着复杂的影响。转录和蛋白合成都可以分为起始、延伸和终止等时期,每个时期都可以被独立调节(参见第 29 章转录,第 23 章蛋白质合成)。在细菌中,转录蛋白合成和 RNA 降解在相同细胞空间中同时进行,基因表达的不同阶段之间可能存在着交叉调节(参见衰减、反向调节)。在真核生物中,mRNA 前体在细胞核内加工,然后运输到胞浆中进行翻译,加工和转运过程都可以被调节(参见第 27 章 RNA 加工)。蛋白合成后,蛋白可能被共价修饰及与其他分子发生非共价相互作用,以此调节它们的活性(参见第 22 章蛋白质:结构、功能与进化)。一个很重要但很少被重视的基因调节阶段是 RNA 和蛋白质的降解速率。蛋白质的降解被认为是基因表达的最终阶段,在细胞周期(参阅)中尤为重要。

表 9.2 基因表达的主要阶段和调节机制

基因表达的阶段	涉及的调节过程
基因的准备	去膜(病毒) 基因组模板链的合成(病毒) 染色质结构的改变(真核生物) DNA 拓扑和构象性质的改变 DNA 甲基化状态的变化 基因组重排 基因扩增
转录	启动子的使用 活性转录复合物的形成 启动子逃避 延伸或暂停 终止或抗终止(原核生物) 衰减(原核生物)
RNA 加工(真核生物)	加帽 多聚腺苷酸化 内含子剪接 RNA 编辑
RNA 转运	mRNA 从细胞核运输到胞浆(真核生物) RNA 定位
RNA 去除	RNA 降解 反向调节(原核生物)
蛋白合成	核糖体结合/蛋白合成起始 密码子的使用 移框 通读/硒代半胱氨酸插入
蛋白修饰	残基的化学修饰 多蛋白的剪切 蛋白内含子的剪接 四级结构的选取 与调节蛋白的相互作用
蛋白定位	细胞空间定位 分泌
蛋白去除	蛋白的丢失和降解

在其他章中对每个阶段均都有较详细的阐述。

9.2 基因调节

基因调控的原理 **基因调控**是指细胞用来控制各基因产物产出量的机制。即使最简单的细胞也有数以百计的基因,基因产物既不是同时都需要,也不需要相同的量。然而基因调节不仅是控制在需要时表达单个基因的机制,同时还要避免不同基因产物之间的无意义的相互作用。没有基因调节就没有多细胞的**后生动物**(多细胞生物)的分化,即没有细胞类型的区别。调节可以是**组成型的**,也就是表达速率根据细胞一般需求保持一定,也可以是**可调的**,表达的速率适合不同环境下细胞的需求,或是在后生动物中反映了不同细胞类型的特殊属性。理论上,基因调节可以在基因表达的任何阶段进行(表 9.2),但实际上,主要发生在转录起始阶段。因为这是表达的初始阶段,可以避免那些不需要的转录所造成的资源浪费。

顺式作用元件和反式作用因子 核酸的调控过程可以是顺式的或反式的(图 9.1)。在 DNA 中,**顺式作用元件**是它调节邻近基因的核酸序列(即同一染色体上的基因;顺式构型的基因),对其他染色体上的基因(反式构型)没有影响。DNA 上的顺式作用元件可调节转录过程,转录成的 mRNA 可以调节 RNA 加工,转换和蛋白合成。顺式作用元件的突变是顺式显性(*cis*-dominant)的,因为它们只影响邻近的基因,而且不能被同一细胞内反式构型的另一份野生型的拷贝补偿(参见补偿)。

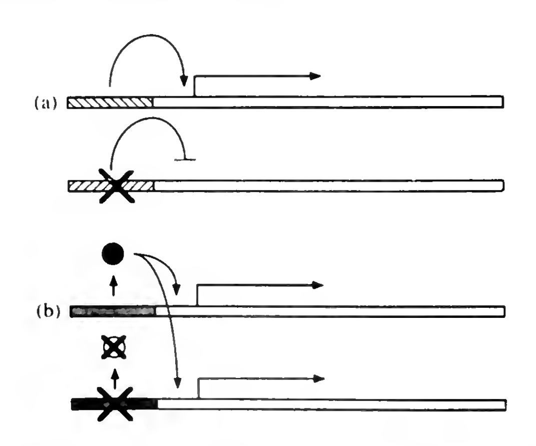


图 9.1 顺式作用元件和反式作用因子。突变以叉来表示。(a) 顺式作用元件的突变是顺式构型显性的,因为它们不能被同一细胞内另一份野生型的拷贝补偿。(b)突变的反式作用因子一般是隐性的,因为同一细胞中的另一份野生型拷贝可以补偿突变基因的功能。偶尔突变因子(以白色圈表示)会阻断野生型因子(黑色圈表示)的功能,产生反式显性突变。

反式作用因子包括各种可扩散分子(经常是蛋白质,有时是 RNA),可调节与它们接触的基因的表达。调节转录的反式作用因子称为转录因子,另一些反式作用因子调控 RNA 加工和蛋白合成。这些因子的突变(编码这些因子的基因突变)一般是隐性的,因为同一细胞中的另一份野生型拷贝可以提供这类可扩散因子的大部分。一类特殊的反式显性突变使突变的因子通过与目标序列的不可逆结合或形成没有活性的多聚体,干扰野生

型的功能(参见显性失活突变)。

偶尔也有顺式作用因子和反式作用元件。顺式作用蛋白立即与其编码 DNA 链作用,形成不可逆结合。这样的因子只存在于转录和蛋白合成紧密相连的原核生物中,如 *Tn5* 编码的转座酶(见第 13 章可移动的遗传因子)和噬菌体 ϕX174 编码的 A 蛋白(见病毒一章)。反式作用元件在称为**反式感应**(*trans*-sensing)的现象中也会遇到,此现象是一条染色体上的调节元件可以影响另一条同源配对的染色体上反式构型基因的表达。这些过程中较为常见的是联合调控(transvection),由于转录因子跨越在两条染色体上,在一条染色体上的启动子可以控制配对染色体的基因表达。在果蝇中有丝分裂期细胞中同源染色体配对时就出现这样**联会依赖性的过程**(synapsis-dependent processes)。在其他真核生物包括人类中当同源染色体连结时,偶尔也有发生这种情况(另参见同源依赖性沉默)。

基因调控模式　　基因调控分为正、负调控及诱导和抑制。

如果基因处在**正调控下**,特定调控因子的存在增加(**上调**)基因的表达,当调控因子缺乏时使表达下降(**下调**),这样的因子为**激活物**(activator)。在**负调控**中,特殊调控因子的存在使基因表达减少,而当其缺乏则表达增加;这类因子称为**阻遏物**(repressor)。

正常情况下基因的表达往往显示了基因调节的模式。**可诱导的基因**以低水平表达,不是因为被阻遏(负调节)就是缺乏激活物(正调节)。去阻遏/阻遏物或激活(激活物的提供)使基因表达被诱导。**可阻遏基因**往往是活化的,由于阻遏物缺乏(负调控)或存在激活物(正调控),阻遏通过去除激活物或提供阻遏物使基因表达下降。

这四种基因表达的模式经常可以在转录调节中发现,同时这些概念适用于基因表达的任何阶段。细菌中调控的四种模式在转录水平可能是同样普遍的,而在真核生物中,正的可诱导调节占主要地位,因为基本的转录体系在遗传上是不稳定的,基因的转录是被阻遏的,除非有激活因子提供才启动转录(参见第 29 章转录)。

基因调节的范围　　**全局调控**(global regulation)、**广泛区域调控**(wide domain regulation)、**共调节**和**局部区域调控**(narrow domain regulation)经常被用来描述调节机制的适用范围。全局调节机制同时控制很多基因,反映了细胞的基本变化。真核生物中的全局转录调控可能由染色质结构控制(如在有丝分裂时转录停止)。同样,在蛋白合成过程中通过阻断 mRNA 前体加工或调节翻译延伸因子的合成来进行全局调控。广泛和局部区域调控机制是指对一种基因或一组基因的调控,通过对单个启动子和增强子的转录激活来实现。广泛区域调节的例子包括饥饿对酵母氨基酸合成基因的共调节、雌性果蝇 X 染色体基因的剂量补偿(参阅),大肠杆菌的 SOS 反应。局部区域控制包括单个基因的调控和一小组基因的协同调控,如乳糖对大肠杆菌乳糖操纵子 *lac* 三个基因的调节。

9.3　真核和原核生物的基因表达

基因调节策略的比较　　尽管原核和真核生物都遵循中心法则,但它们的基本生物学属性有着各种差异,反映在它们基因表达和调节的机制和策略上存在差异。

(1)真核生物细胞核　　细菌没有细胞核,基因表达的不同阶段都在细胞的同一空间进行。转录、蛋白合成和 RNA 更新是紧密相连的,允许像衰减和反向调节(参阅)这样在

真核生物中不存在的调节机制。真核生物细胞核的出现将转录和蛋白合成分在不同空间和时间进行。mRNA 转运出细胞核的过程,增加了调控层次。

(2)顺式调控的协调　在细菌中,类同功能的基因经常排列成基因簇,受到同样的转录调节,称为操纵子(参阅)。多基因转录为多顺反子 mRNA,这使基于操纵子内基因的位置进行基因调节(如转录的抗终止和蛋白合成时核糖体结合的顺序依赖性)成为可能。基因排列相近意味着可以应用如逆向转录(countertranscription,参阅)的策略,逆向转录是指附近的启动子用来调节基因的表达。操纵子的结构在真核生物系统中几乎完全被摒弃,因为在真核生物中核糖体只能与 mRNA 经过修饰的 5′端结合,而不能与中间位置结合(参见内部核糖体进入位点)。另外,真核生物的基因分布比原核生物松散。这两个因素都阻止真核生物协同的顺式转录,尽管共享远端的调节元件时转录的顺式共调节还是可能的。

(3)内含子　高等真核生物的基因经常被一个或多个内含子(参阅)打断,内含子所含信息并不出现在最终的基因产物中。一些低等真核生物内含子很少出现(如酵母),在细菌中则几乎不出现内含子。内含子通过真核生物 RNA 剪接(参阅)而被去除,这样在转录和蛋白合成中增加了一个调节阶段,并能有选择地使用基因的信息。细菌基因一般只编码一种基因产物而很多真核生物基因可以编码多个相互重叠的产物。

(4)染色质　在真核和原核生物中 DNA 结构对基因表达的调节都是全局的调控机制,因为 DNA 结构可以影响转录调节因子与 DNA 的结合。在真核生物中,DNA 包装成被称为染色质(参阅)的高度有序核蛋白复合物,它的结构在有潜在转录活性的形式和转录抑制的凝聚状态之间转换。染色质结构对基因活性的调节使大部分基因完全被抑制或激活,这对含有庞大基因组的生物可能是有利的。

(5)RNA 聚合酶活性　在细菌中,RNA 聚合酶全酶有一亚单位——σ 因子,促进 RNA 聚合酶对启动子识别和结合。如果存在合适的启动子,不需其他调控方式就能进行有效的转录。在真核生物中,RNA 聚合酶本身不能识别启动子,需要一系列基本转录因子促进装配和模板的进行性。起始复合物不稳定且只有很弱的转录活性。有效的转录还需要组建成型的正调节因子来稳定起始复合物。

(6)转录调节　在细菌中,转录的调节主要是由直接与 RNA 聚合酶作用的因子优先介导的,这些因子经常受别构调节。这些以及 mRNA 的高更新率有助于对不断变化的环境条件作出快速反应。在真核生物中,转录调节常依赖于结合在远处基本启动子上的因子,这种相互作用由环出的间隔 DNA 介导。这样允许基因竞争远处的增强子和其他调节元件的协同顺式调节发生(参见增强子竞争、基因座控制区域)。在细菌中,对已经存在的转录因子的别构改变可以介导对细胞外诱导事件的迅速反应。在真核生物中,转录因子翻译后加工也是一种普遍的调节策略,而重新合成新的转录因子用来调节相当缓慢和长期的反应,如发育过程中的分化(参见第 6 章发育,分子机制)。

(徐　沁译)

第10章
细菌中的基因转移

基本概念和定义

- **基因转移**(gene transfer)是指外源(经常指另一个细胞)的遗传信息进入一个细胞的现象。这种过程在细菌和真核生物中都会自然发生,称为**水平或横向遗传传递**(horizontal or lateral genetic transmission),区别于称为**纵向遗传传递**(vertical genetic transmission)的亲代到子代遗传信息的传递。
- 种内基因转移有利于无性繁殖物种中的遗传混合,它模仿了有性生殖的效果。这样的**准性交换机制**(parasexual exchange mechanisms)可用来对原核基因组作图,类似与真核减数分裂作图(参阅)。种间基因转移也可能发生,是转基因(参阅)的自然机制。种间基因转移是重要的进化过程,与一些最基本的进化事件有关(如真核细胞器的内共生体起源),也与细菌和真核生物间特异的相互作用有关[如根瘤农杆菌(*Agrobacterium tumorfaciens*)的肿瘤诱导;参见 Ti 质粒]。
- 被转移信息的来源称为**供体**(donor),被转移的遗传信息称为**外基因子**(exogenote,外源基因组,一般是供体基因组的一个片段)。基因转移的目标是**受体**(recipient),它具有**内基因子**(endogenote,内源基因组)。如果外基因子与内基因子的一部分同源,基因转移会使受体细胞成为部分二倍体(**部分合子**,merozygote),这种情况下可能发生重组,包括等位基因的替换(**标记基因交换**)。
- 细菌中基因转移有四种主要的机制:细胞融合、接合、转化和转导(表10.1)。所有基因转移的机制都包括将DNA导入受体和一个决定外基因子命运的分离期。基因转移有四种结果:拒绝(降解)、保留(也就是处于细胞质中)、置换(与内基因子重组)和叠加(整合到内基因子中)。置换只有在外基因子与部分内基因子同源时才能发生。
- 人工基因转移用于基因的克隆与表达、菌株构建、定靶突变、产生转基因生物和基因表达分析。在这些情况下,外基因子是重组DNA分子,进一步讨论见重组DNA。

10.1 接合

接合的要求　细菌接合包括在细胞发生物理接触时遗传信息从一个细胞转移到另一个细 胞。供体细胞被人为地定义为**雄性**,而受体细胞为**雌性**,基因转移后,受体可被称为**转**

表 10.1　细菌基因转移的四种基本机制

基因转移机制	定义特征
接合	DNA 从细胞到细胞直接转移，经常通过一个特定的管道。 接合是质粒转移机制进化而来，需要许多特殊的功能，往往由质粒提供。
转化	细胞从周围介质中吸收裸露的 DNA。在细菌中转化是指从周围介质中吸收裸露的染色体或质粒 DNA。对裸露的病毒基因组 DNA/RNA 的吸收称为**转染**(transfection)。极少种属能自发有转化能力，在一些耐大肠杆菌素的种属中通过不同的化学处理，能被诱导产生转化的人工感受态(参见第 24 章重组 DNA)。 在动物细胞中，从周围介质中吸收任何裸露 DNA 都称为**转染**，尽管转化这个名称可以用来特指因此带来的表型改变。酵母和植物细胞中导入裸露的 DNA 可以称为转染或转化。
转导	通过病毒颗粒介导将染色体或质粒 DNA 转移到细胞中。DNA 可能取代病毒基因组被包装在噬菌粒中(普遍性转导)或共价地加入到病毒基因组中(特异性转导)。
细胞融合	发生在有些细菌种属中的由细胞质膜的融合介导基因转移(如链霉菌属)。

注意转化和转染两个术语在细菌和真核生物中的不同使用方法。

接合子(transconjugant)。通过接合而转移 DNA 的能力是由**接合质粒**(conjugative plasmid)提供(或很少情况下是**接合转座子**提供)，他们是**自身可传送**的元件，编码所有通过接合将自身拷贝转移到其他细胞所需要的功能。如果接合质粒能促进染色体基因的转移，它就称为**性因子**(sex factor)。

F 质粒(F 因子、致育因子)是大肠杆菌的一种接合质粒，它除了致育(刺激接合转移的能力)，不赋予其宿主其他表型。F 因子是第一个发现的，接合系统研究得也最为清楚(见文框10.1)。由于与其他接合质粒不同，它的接合基因是组成型去阻遏的，因而在自发状态下，F 启动从供体(雄性，F^+)细胞到受体(雌性，F^-)的高效自身转移。在雄性和雌性细胞的混合群体中，大多数雌性细胞接受了 F 的拷贝成为雄性，但在很高的雄:雌的比例时，雌性可能由于接合转移被杀死，这一现象称为**致死接合** (lethal zygosis)。**R 质粒(R 因子，抗药因子)**是携带抗生素抗性标记的接合质粒。因为转移基因被抑制，R 质粒促进接合的效率不像 F 因子那么高；它们通过反式抑制 F 转移基因，对 F 显示**致育抑制**。这可发生在 F 系统中，因为 F 型质粒在两个不相容组中扩散(参阅)。

质粒 DNA 的转移和命运　接合需要细胞间的接触，这样 DNA 才能从供体转移到受体。在革兰氏阴性细菌中(如大肠杆菌)，这样的接触包括质粒编码的蛋白管道(**伞毛、性伞毛、接合管；复伞毛**)，它从供体延伸到受体细胞表面将接合细胞拉在一起。伞毛是成功接合必需的，可表明接合在何种条件下发生。一旦接触成功发生，就释放启动 DNA 移动的信号。在革兰氏阳性细菌中(如链霉菌属)，接合不需要菌毛而是细胞间直接接触。

一旦建立了连接，接合质粒的 DNA 就进入可移动状态(为转移做准备)。在 F 质粒的情况下，这包括通过在**转移起始点**(origin of tranfer)*oriT* 处导入缺口，解开单链 DNA，并传送入伞毛中。DNA 的 5′端最先被吸收，因此 *oriT* 缺口的 5′位置先进入受体细胞。在受体细胞中，单链 DNA 作为模板产生双链分子，再重新环化，这种过程称为**独立复制**(repliconation)。在供体细胞中，保留的单链也作为模板替代被转移的链(**供体接合 DNA 合成**)。因此接合是半保留的过程(图 10.1)。

染色体 DNA 的转移和命运　接合质粒不仅自身可以转移，而且还促进自身不能转移的染色体或其他质粒 DNA 的接合转移。如果外源 DNA 与接合质粒共价接合，这种情况可以顺式作用形式发生(**顺式移动**或**共转导**)，如果另一质粒是可移动的，也可以以反式发生(也就是它可以对反式作用的移动因子如核酸内切酶 TraY，产生反应，但本身不编码这样的因子；**反式移动**或**给体作用**)。

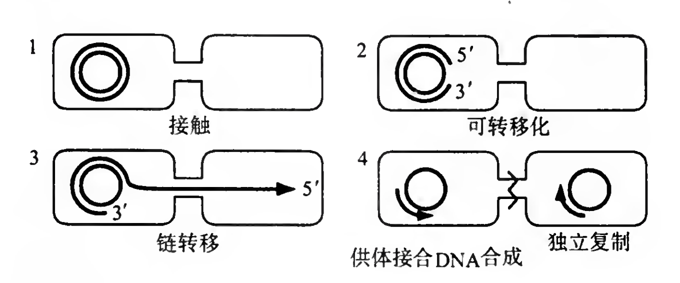


图 10.1　质粒 DNA 接合转移的步骤。

当接合转移涉及染色体基因的移动和转移，此过程将促进准性交换，称为**性导**(sexduction)(或在 F 质粒存在时为 **F 因子转导**)，可用于对原核基因作图(文框 10.2)。F-质粒可以通过 IS 因子之间(参阅)的同源重组整合到宿主的染色体中。一旦整合，它就可以顺式移动它所共价连接的染色体 DNA，使染色体 DNA 从一个体转移到另个个体中。受体细胞对转导的基因而言是二倍体，如果供体和受体细胞是不同的基因型，随着外基因子和内基因子的同源重组，标记基因产生互换。整合有 F 质粒的细菌菌株被称为 **Hfr 菌株**，因为它们促进了在供体和受体起始点的染色体标记的*高频重组*(high frequency of recombination)。当供体菌株具有自发的 F 质粒时，只能观察到低频率的标记交换，被认为存在少量的游离 F 携带者所致。自发 F 质粒的转移总是使 F^-(雌性)受体细胞变成 F^+(雄性)，而在 Hfr 菌株中不会发生这种情况。其中原因反映了转移区域内的 *oriT* 位置。因为 *oriT* 位于转移区域的 5′边界(见文框 10.1)，紧连 *oriT* 5′的基因座最先进入受体细胞，转移基因经常最后进入受体细胞。游离 F 质粒因为经常整个质粒转移而不受影响；然而如果质粒是整合到细菌染色体中，接合转移几乎总是在整个染色体被转移前被打断，这样使受体细胞保持 F^-。

游离的 F 质粒是偶尔从宿主基因组中不精确切离而产生的，它携带部分染色体 DNA(参见*误切质粒*)。增大后的质粒，现在称为 F′质粒，能够转移到 F^- 细菌中，使它们变成 F^+，且对额外基因是二倍体。F′质粒文库[**养粒**(fosmid)文库，参见*粘粒*]可以用来分离包含整个大肠杆菌染色体的区域，这样的质粒可以有很多用途：在重组缺陷的宿主中，部分二倍体可以用于研究*反式补偿*(参阅)的突变效应；它是精细分析乳糖操纵子的方法；在容易重组的受体中，在染色体和质粒携带基因间可以发生标记的交换，类似于*特异性转导*(参阅)机制的性导过程可用于菌株构建。F 质粒和 F′质粒的性导过程在图 10.2 中进行了比较。

10.2　转化

转化的要求　在细菌中，转化是从周围介质吸收裸露DNA，使受体细胞，即**转化体**

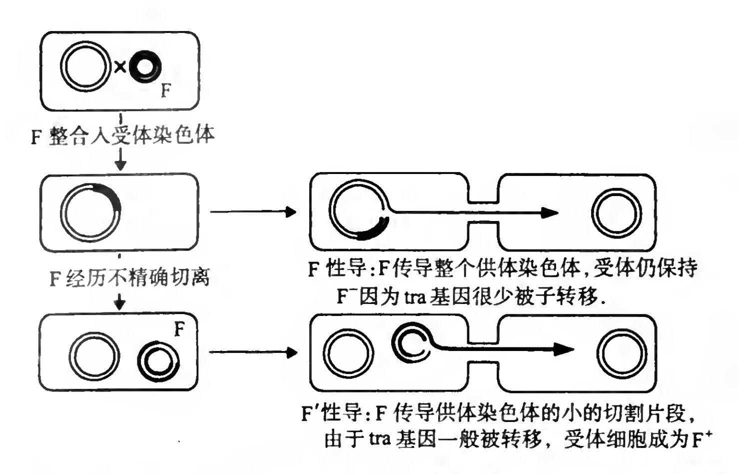


图 10.2　性导:接合时染色体 DNA 的转移。通过传导其整合入的染色体或传导由不精确切离所捕获的染色体基因来介导性转导。

(trasformant)基因型发生改变。很多细菌中自然发生转化(如杆菌、链霉菌属、嗜血杆菌),尽管**感受态**(competence)(有能力吸收外源 DNA 的状态)经常是瞬时的,与特定的生理状态有关,并需要表达特定的**感受态因子**(competence factors)。其他细菌种属,包括大肠杆菌,自然转化无法发生,但可以通过人工诱导产生感受态,使其能够吸收 DNA;促进大肠杆菌在分子克隆中的使用(见第 24 章重组 DNA)。

DNA 的转移和命运　　在自然的感受态细胞中,供体 DNA 先与表面受体可逆结合。在有些细菌中(如杆菌),DNA 通过剪切和降解,最后只有一条链被内化。在其他菌株(如嗜血流感杆菌)中两条链都进入细胞。人工转化大肠杆菌也可使 DNA 被完整地内化,可能是因为处理过程中增加了细胞膜对 DNA 通透性。一般,DNA 的结合和内化是非特异性的,尽管在有些例子中,如嗜血流感杆菌只吸收有特定 **DNA 吸收位点**的 DNA,DNA 吸收位点是在嗜血流感杆菌基因组出现频率很高的顺序,这样保证细胞只转化从同一种属来的 DNA。

如果转化的 DNA 是质粒,它在受体细胞中可以作为自主复制子存在。线性化染色体 DNA 可以与宿主染色体重组,产生标记交换。这两种情况都产生**稳定的或永久的转化**。另外转化的 DNA 也可能被降解,在这种情况下,它所提供的任何特性是短期的(**瞬时转化**)。偶尔转化的 DNA 可以通过非法的末端整合连接到宿主染色体中,尽管由于富含末端连接修复酶,在真核细胞中线性化 DNA 的导入是更常见的现象(参见转染、转基因、非法重组)。

10.3　转导

通过普遍性转导和特异性转导转移 DNA　　**转导**(transduction)是通过病毒颗粒将细胞内基因从供体转移到受体细胞的过程,受体在转移后称为转导体(transductant)(参见信号传导)。自然的转导可以两种方式发生(表10.2)。在广泛性转导中,染色体或质粒

表 10.2 广泛转导和特导转导的特性

	广泛性转导	特异性转导
转导颗粒的成分	只有宿主 DNA,理论上任何顺序都可以	与噬菌体 DNA 共价连接的宿主 DNA。噬菌体经常是缺陷的。宿主 DNA 限制在原噬菌体插入位置的两侧
机制	在裂解循环中通过错误的对宿主 DNA 包装进入衣壳形成颗粒	在原噬菌体异常剪切后形成颗粒
分辨力	**完全转导**:与宿主基因组同源重组——转导体是单倍体	**置换转导**:与宿主基因组同源重组——转导体是单倍体
	流产转导:供体 DNA 保留在细胞质中,不复制 - 转导体是部分二倍体,一群细胞中只有一个细胞包含转导的基因	**插入转导**:由转导噬菌体溶源化——转导体是部分二倍体
要求	低毒性(噬菌体在包装前不能破坏宿主 DNA),包装 DNA 是非特异性的,如通过*头满装机制*(参阅)	必须是温和性噬菌体,也就是必须整合到宿主基因组中
宿主重组体系	完全转导需要细菌 *rec* 系统,完全:流产转导的比例随供体 DNA 双链断裂增加而增加	置换转导需要细菌 *rec* 系统。加入转导依赖于缺陷噬菌体性质,不一定要求 *rec* 系统
举例	大肠杆菌的噬菌体 P1;沙门氏菌(*S. typhimurium*)的噬菌体 P22	大肠杆菌的 λ 噬菌体;枯草杆菌的噬菌体 SPβ

DNA偶尔替代噬菌体基因组被包装到噬菌体头部。因为感染是由噬菌体颗粒而不是所携带的核苷酸提供的能力,所以这是一种有效的细胞间基因转移的机制,理论上染色体的任何区域都可以被转导。在**特异转导**(**或限制性转导**)中,对原噬菌体不精确的切离使原噬菌体两侧的宿主 DNA 被移入和包装。因此转导的 DNA 是共价地与噬菌体基因组相连接,能够转导的基因被限制在原噬菌体插入位置的两侧。在真核生物中可以观察到特异性转导而不是普遍性转导(参见*急性转化逆病毒*)。

病毒基因组可以用来作为*克隆载体*(参阅),通过将重组病毒包装进入衣壳,将克隆基因转移到克隆宿主中,这个过程可被认为是人工转导;使用重组噬菌体 λ 是人工特异性转导,因为克隆的 DNA 共价地与 λ 基因组连接。相反,用噬菌粒载体克隆更像普遍性转导,因为根本没有使用 λ 基因组(见第 24 章重组 DNA)。

大多数温和的噬菌体可以作为**特异性转导噬菌体**,但极少噬菌体天然就具有普遍性转导的能力——这是因为**普遍性转导噬菌体**必须不破坏宿主 DNA,并且使用不需要特定噬菌体顺序的包装机制,也就是*头满装机制*(见第 30 章病毒)。很多不能作为普遍性转导的野生型噬菌体通过特殊突变可以补偿得到这种能力(如 λ 和 T7 噬菌体)。

普遍性转导噬菌体　　主要的普遍性转导噬菌体是大肠杆菌噬菌体 P1 和沙门氏菌(*S. typhimurium*)的噬菌体 P22。它们的基因组是*环形排列和末端冗余的*(参阅),这是多联体基因组 DNA 包装的结果,环形排列是由从同样的多联体中剪切基因组亚单位的结果,末端冗余是每个末端都填补了比一个基因组大的固定长度基因组 DNA,即*头满装机制*的结果(参阅)。宿主 DNA 替代噬菌体基因组进入头部形成**转导颗粒**。理论上,宿主染色体的任何部分都可以相同的效率被包装和转导。但实际上,不同的标记以不同的频率转

导,有三个数量级的变化。在 P22 转导中这种效应特别明显,这反映了在 P22 基因组中,组成染色体**包装位点**(*pac*)的特定染色体区域的包装倾向。已获得了包装专一性缺陷的突变噬菌体株,它们以类似的效率转导所有的标记。

普遍性转导 DNA 的命运 外源 DNA 通过病毒颗粒进入受体细胞后可能有多种结果。如果转导的 DNA 是质粒,它可以作为自主复制子在细胞质中存在;大的质粒在转导后变小(**转导短化**),这一现象反映了自发缺失突变的包装效率更高。转导染色体 DNA 的线性片断能与受体基因组的同源区域相联会进行同源重组(参阅)和标记交换;这称为**完全转导**(complete transduction)。保留在细胞质中的线性 DNA 可能被降解或在细胞质中作为脱氧核糖体的一部分稳定下来;这称为**流产转导**(abortive trans duction)。转导的基因可以表达,但由于缺乏起点它们不能复制,合成的蛋白会被不断增加的群体所迅速稀释。因此,如果携带一个营养缺陷(参阅)突变的细胞用对应的野生型等位基因转导,一些完全转导体会在选择性培养基中正常生长,而流产转导生长会非常缓慢,产生小的菌落——这是因为转导 DNA 可以产生细胞生长所需要的酶,但不能在细胞分裂时扩增它仅能遗传给一个子代细胞。因此尽管每一子代细胞中可能有足够的酶维持几代细胞的生长,但最终被稀释和降解,所以除了携带片断的细胞,其他细胞停止生长。

利用 λ 噬菌体的特异性转导 特异性转导由原噬菌体的异常切离产生(见第 30 章病毒),使原噬菌体插入位点两侧的宿主基因被转导。很多温和的噬菌体有特殊的插入位点。噬菌体 Mu 是一例外,因为它通过较小靶位点倾向性的重复转座复制。Mu 可以作为普遍性转导噬菌体起作用,Mu 的缺失衍生体可以作为特异性转导噬菌体起作用,这个过程称为(缺陷)**Mu 噬菌体转导**(mini-Muduction)。在 λ 噬菌体的情况中,特异性转导颗粒携带 *gal* 或*bio* 基因座(在 λ 噬菌体插入位置 *aatB* 两侧)。转导是以基因组另一端的噬菌体基因丢失为代价的,即产生一个**缺陷噬菌体**(缺乏关键基因的噬菌体,只有在通过称为**辅助噬菌体**的野生型噬菌体反式提供缺失的功能时才能感染宿主细胞)。这样异常的剪切发生频率不高,用从 bio^+ 宿主产生的裂解液培养感染 bio^- 受体细胞,使少数细胞被感染,然后它们通过 bio^+ 的特异性转导颗粒(λbio^+)被转导。这样的裂解液称为**低频转导裂解液**[low-frequency transducing(LFT) lysates]。如果偶尔 λ 噬菌体整合在与 *attB* 以外的位置,其他基因也能被转导。

特异性转导 DNA 的命运 如果 λbio^+ 感染第二个 λbio^- 的宿主,可能有两种结果(图 10.3)。转导的基因与染色体基因座重组,有利于标记交换。这是**置换转导**,转导体是 bio^+,但保留了单倍体的 *bio* 基因座。另一种情况是特异性转导颗粒中的 DNA 可能整合进入宿主基因组中(**插入转导**)。这只能在有辅助噬菌体的同源重组情况下发生,因为特异性转导颗粒的 DNA 对位置专一性重组缺乏有效的供体位点。因为在 LFT 裂解液中有过量的野生型噬菌体,受转导噬菌体感染的受体细胞同时也感染了野生型颗粒。两种噬菌体的整合产生了**双重溶原**(double lysogen),宿主对 *bio* 基因座是二倍体。这样的转导被称为**溶原部分合子**(lysogenic merozygote),通过原噬菌体的整合有一个额外 *bio* 拷贝。后来的双重溶原诱导产生包含等量的野生型噬菌体和特异性转导噬菌体的裂解液,它可以高频率转导另一群 bio^- 的受体细胞[**高频率转导(HFT)裂解液**]。

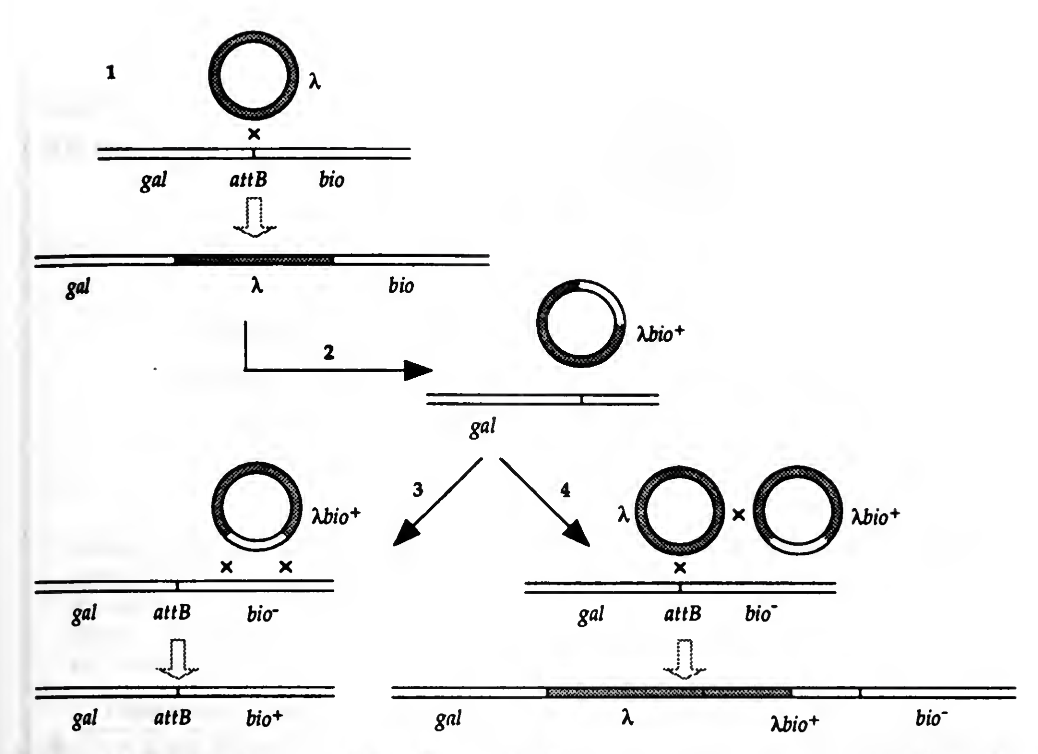
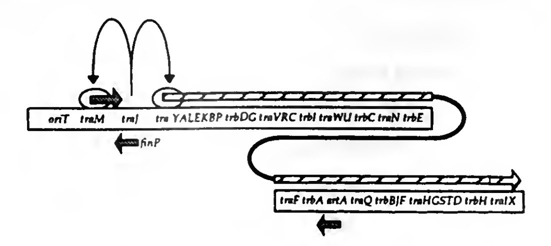


图 10.3 λ噬菌体的特异转导。(1) λ噬菌体整合在大肠杆菌 *gal* 和 *bio* 间的 *attB* 位点。(2)异常切离产生携带 *bio* 基因的特异性转导颗粒 λ*bio*$^+$,再次对 *bio*$^-$ 宿主感染造成(3)重组引起的置换转导或(4)整合引起的插入转导,后者产生一 λ:λ*bio*$^+$ 双重溶原,可以产生高频率转导的裂解液。在 *gal* 基因座的情况是类似的。

文框 10.1 F 质粒上的转移基因

tra 区域和它的调节 大约三分之一的 F 质粒包含一个**转移操纵子**(transfer operon),约 35 个基因,被 *traJ* 基因产物正调节。在操纵子的 5′端有 3 个基因座:转移起点 *oriT*,受 TraJ 控制的独立转移基因 *traM* 和 *traJ* 本身。这 3 个基因座和下游的操纵子组成了**转移区域**(transfer region),携带了所有自身转移所需要的基因。在其他接合质粒中,*traJ* 的表达被两个基因 *finO* 和 *finP* 的作用所抑制,后者编码小 RNA 分子,与 FinO 一起结合在 *traJ* 的前导顺序上,抑制其表达[反义 RNA 是质粒(参阅)基因调节的常用策略]。在 F 质粒中,*finO* 基因被 IS 因子(参阅)的插入打断,转移基因组成性表达。图中是与大肠杆菌染色体重组的 IS 因子。



F 质粒的 *tra* 区域包含(从 5′到 3′方向)*oriT*、*traM*、*traJ* 和 *tra* 操纵子。*tra* 操纵子是连续的;如上所示,*trbE* 和 *trbF* 之间没有物理断裂。

tra 操纵子的基因　大约 20 个转移基因的功能已经知道。其他如产物大小、及它在很多情况中的亚细胞定位和可能相互作用配偶体也已经得知，但精确的功能还有待于进一步确定。***traI*** 基因座通过嵌套翻译编码 **TraI** 和 **TraI***（以前是 **TraZ**）两种产物，**TraI*** 的功能还不清楚。

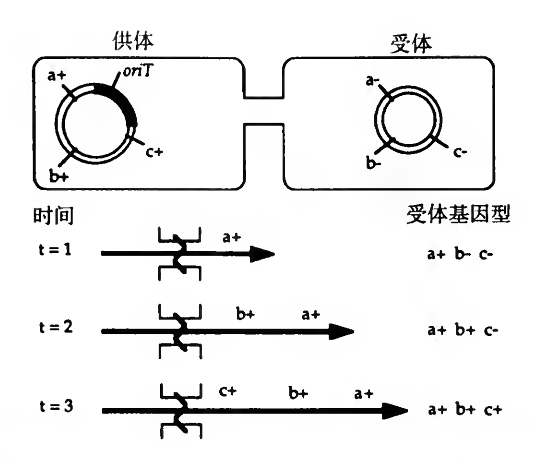
基因	基因产物功能
traM	衣壳蛋白，可能涉及细胞-细胞识别
traJ	是 *traM* 和 *tra* 操纵子转录正调节因子
traY	内切酶，在 *oriT* 处产生 DNA 缺口
traA	编码性伞毛的主要亚基菌毛蛋白
traL、*traE*、*traK*、*traB*、*traV*、*traC*、*traW*	性伞毛装配
traU、*trbC*、*traF*、*traH*、*traG*	性伞毛装配与交配集合物稳定化
traN	性伞毛交配集合物稳定化
traS、*traT*	表面排斥[1]
traD	接合时 DNA 转运
traI	解旋酶，转移所需的 DNA 解旋。也有内切酶活性

1 **表面排斥**（surface exclusion）是阻止供体细菌与其他携带相同接合体系的供体细菌接合的现象。

文框 10.2　细菌连锁作图

细菌的遗传作图　细菌中准性交换被用来进行基因作图，这和真核生物中利用减数分裂交换作图的方法相同（见重组、基因组和作图）。细菌连锁作图的策略十分精致，并提供了几种原核生物基因组的详细图谱，但它现在逐渐被快速精确确定基因组 DNA 克隆的序列和高产量的测序策略等先进物理方法所取代（见第 24 章重组 DNA，第 12 章基因组和作图）。下面讨论的作图方法证明了遗传分析的力量，而不仅仅出于历史的兴趣。

通过性导作图　如果野生型的大肠杆菌 Hfr 株与携带很多突变的 F^- 细胞混合，野生型的等位基因从供体传递到受体，标记交换也可能发生，所以转接合子使一个或多个突变基因座变成野生型。因为染色体以线性形式转移到受体细胞，以 *oriT* 的 5′端为 DNA 的开始，位于供体染色体接近 *oriT* 5′端的标记基因最先进入细胞。在任何细胞群中，一对接合细胞可以随机分离；这样处在离 *oriT* 较远的标记基因进入受体细胞的机会要少一些。这样建立了一个**转移梯度**（gradient of transfer），在这个梯度中，接近 *oriT* 的 5′边界的标记比远处的标记转移频率高，发生标记交换的频率也要高。转移梯度可用来确定基因座顺序。如果由 Hfr 和 F^- 细胞起始接合，然后将样品取出，在不同的时间振荡来阻断接合对，在阻断较晚的样品中有更多的染色体被转移，*oriT* 较远处的标记可进行重组。这样的**中断杂交**（interrupted mating）实验可以用来确定基因在染色体上的顺序，并估测它们之间的图距，以分钟来表示的（见图）。



中断杂交作图的原理：一 Hfr 供体以线性方式转导染色体进入受体细胞。经过时间 t，接合被振荡打断(折线表示)。*oriT* 5′侧附近位置的基因座会先被转移，所以在时间 $t=1$ 时，只有标记 a 被转移和交换，产生表型为 a+b−c− 的转接合体。在时间 $t=2$，标记 b 被转移和交换产生表型为 a+b+c− 的接合体。在时间 $t=3$ 时，标记 c 被转移，产生表型为 a+b+c+ 的转接合体。这样建立了基因顺序，用每个标记转移所需时间所反映的距离构建一个粗略的图谱。

通过共转导和共转化作图 中断杂交可测得基因座的顺序和对它们之间的距离有一个概念。但它不能应用于精细作图，因为紧密连锁的标记转移的分离时间只差几秒。为了更精确，根据不同的细菌菌株采用了**共转导或共转化作图**。普遍性转导可以频繁发生(野生型 P1 和 P22 噬菌体有 1%～5% 为转导颗粒)，但任何特定基因座的成功转导以很低的频率发生(大约 10^{-6})，而**共转导**(两个基因座同时转导)被作为两个基因在同一 DNA 片段上连锁的证据。这样精细基因作图可以通过测定一对供体染色体上标记的**共转导频率**(cotransduction frequency)来进行，高的共转导频率表明强的连锁。在枯草杆菌中，通过**共转化作图**(两个基因座同时转化)是以同样的原理进行的。在一定的 DNA 浓度下(每个细胞只吸收一个 DNA 片段)，**共转化**是连锁的证据，**共转化频率**(cotransformation frequency)反映了基因座之间的距离。如果一些连锁的基因以上面三种机制中的任何一种转移到受体细胞中，基因座之间的距离可以通过**重组作图**(recombination mapping)估测，也就是通过确定标记基因对之间的重组频率估测物理距离。

(徐　沁 译)

第11章
遗传密码

基本概念和定义

- **遗传信息**(genetic information)指 DNA 中碱基序列。遗传信息也可能存在于染色体物理结构中,与碱基序列不直接相关,这称为**外遗传信息**(epigenetic information)(参见染色质、DNA 甲基化、遗传外调节和核酸结构各章)。
- 遗传信息要以多肽序列形式表达,就必须将核酸的 4 字符语言翻译成多肽序列的 21 字符语言,用核酸中的碱基序列定义特异的多肽的氨基酸序列,也就是说使基因编码蛋白质结构的体系就是**遗传密码**(genetic code)。遗传密码是近乎通用的三联体密码,只有一种阅读方法,但却是高度简并的。
- 翻译通过 tRNA 进行:tRNA 是小分子 RNA,它拥有氨基酸结合位点,可识别 mRNA上密码子的反密码子而扮演了适配器的角色。这样 tRNA 携带正确的氨基酸带到核糖体上进行蛋白合成。
- 蛋白质含有很多不同类型的氨基酸,但遗传密码只对应有 21 种氨基酸,其余由翻译后加工产生(参见蛋白质一章)。

11.1 遗传密码的概论

遗传密码的性质 遗传密码用三个碱基的组合定义一种氨基酸。三碱基的一种组合称为**密码**或**三联体**(triplet),前者适用于 mRNA,后者对基因而言。mRNA 上**有义密码**(sense codon)的线性排列决定了多肽上氨基酸的线性排列。特别的密码子决定多肽链的起始和终止。**起始密码子**(initiator codons)总是编码甲硫氨酸(原核生物和细胞器中是甲酰甲硫氨酸),而**终止密码子**(stop codons)不编码氨基酸,而作为解散核糖体的**释放因子**(release factors)作用的信号。终止密码子也称为**无义密码子**(nonsense codons),因为没有可翻译的意义。无义密码子有三种,分别称为**琥珀**(amber,UAG)、**赭石**(ochre,UAA)和**乳白**(opal,UAG)或**赭色**(umber)。遗传密码的性质列在表 11.1,通用遗传密码在图 11.1 中表示。

密码子由蛋白质合成机器从起始到终止以不重叠和不停顿的方式进行"阅读",即每个核苷酸三联体相对独立且在密码子之间没有空格。这样任何核苷酸顺序都可能以三种方式分成不同的密码子,也就是说有三种**读框**(reading frames)(图 11.2)。蛋白合成所用的阅读框架决定于起始密码子的位置,偶尔在翻译过程中也有读框的改变(参见移框突变,frameshift mutation,**共翻译移框**,cotranslationsl frameshift)。在基因组顺序资料中,潜

在的基因往往是根据**开放阅读框架**(open reading frames, ORFs)而定义,也就是不间断的有义密码子的长顺序,因为不正确的阅读框架往往含有一个或多个终止密码子(这就是为什么移框突变会使基因产物截短的原因)。当一个阅读框架被鉴定而其产物还不清楚,就称为**未鉴定的读框**(unassigned reading frame,URF)。

表 11.1 遗传密码的性质和已知的例外情况

性质	意义	例外
通用性	在所有生物体中密码相同(**通用遗传密码**,universal ngenetic code)。	细胞器和低等真核生物基因组中有例外情况(参见表 11.2)。
不重叠	每个密码子都是独立的单位,密码子以 1-2-3,4-5-6,7-8-9 顺序阅读,而不是 1-2-3,2-3-4,3-4-5。	翻译过程中的移框和绕道(参见**翻译调节**)。
无标点	密码子之间没有空格。	
单一读法	每个密码子只对应一种氨基酸或终止信号。	不常用的起始密码子,硒代半胱氨酸在 UGA 处插入。
简并性	有61 个有义密码子和 21 种氨基酸,每种氨基酸(或终止信号)可以由多个密码子定义。	密码子作为整体而言是简并的,色氨酸和(内部的)甲硫氨酸在通用密码子中只有一个密码子对应。

图 11.1 通用遗传密码子。氨基酸以三字符命名(见文框 22.1)。无义密码子根据名字定义,注意密码子 AUG 和 UGA。

图 11.2 阅读框和在蛋白合成过程中如何选择正确的阅读框。

11.2 翻译

蛋白合成过程中密码子的解译 蛋白合成阶段称为翻译，指对遗传信息的解译，即mRNA中的密码子和所对应的氨基酸相匹配的过程。这一过程需要转移RNA(tRNA)(参阅)，tRNA结构中包括了**反密码子**(anticodon)(与密码子互补的三碱基序列)和一共价结合氨基酸的受体柄。当tRNA与氨基酸结合即被装载，称为**氨酰tRNA**(aminoacyl-tRNA)。氨酰tRNA进入核糖体A位点，反密码子与密码子配对。在这个位置上，氨基酸处于与核糖体肽基转移酶活性位点邻近的位置，后者催化氨基酸从tRNA转移到存在的多肽链上(详见蛋白合成一章)。

翻译的保真度 翻译过程的精确性通过不同**氨酰tRNA合成酶**家族来控制，氨酰tRNA合成酶的功能是负责tRNA与其相应氨基酸的结合。这些酶对底物的专一性是通过特殊识别机制，而不是和任何特殊结构相关(参见核酸结合蛋白)。有多少氨基酸就有多少氨酰tRNA合成酶，每种酶都能识别所有的同工tRNA(参阅)。

转移RNA的装载有两个阶段，称为激活和转移，二者都可以被校正。在**激活**(activation)过程中，氨基酸与氨酰tRNA合成酶相连，产生活化复合物；这过程依赖于ATP结合和水解。很多氨酰tRNA合成酶与氨基酸的结合是一步识别，如果氨基酸与激活位点不匹配就将其抛弃(色氨酸就是以这种方式识别的)。那些在类似底物(如丙氨酸和异亮氨酸)之间选择的氨酰tRNA合成酶可能在识别步骤时两种底物中的任一个都能结合，而在以后的阶段将不正确的氨基酸抛弃。在**转移**过程中，激活的复合物结合tRNA，氨酰基团从酶转移到tRNA的3′端腺嘌呤上，并释放出AMP。根据反应的性质，氨酰tRNA合成酶分为两类：I类酶将氨基酸转移到3′羟基而II类酶将氨基转移到2′羟基端。酶开始与tRNA开始的作用引起构象的改变，如上文所述，如果氨基酸不正确就将其抛弃(**转移前校正**)。氨基酸转移到tRNA后，酶识别产物的外形，如果tRNA装载不正确就水解肽键(**转移后校正**)。转移前后的选择称为**双筛校正**(two sieve proofreading mechanism)。

11.3 密码子的特殊属性

同工tRNAs和摇摆碱基配对 遗传密码表现出两种类型的简并性：**第一和第二位碱基的简并性**(密码子前两位不同碱基可能编码相同的氨基酸)和**第三位碱基的简并性**(密码子不同的第三位碱基可能编码相同的氨基酸)。对应相同氨基酸的密码子称为**密码子家族**(codon family)，其成员称为同义密码子(synonymous codons)。密码子家族规模最大的为六个，最小为一个(图11.1)。

第一和第二位碱基的简并性反映了存在同工**tRNAs**的存在。原核和真核生物都合成大约30种不同种类的tRNA，但对应遗传密码子的只有21种氨基酸，某些tRNA种类必然识别相同的氨基酸。其中类似的tRNA有相同的反密码子，它们在功能上可相互替换。而另一些tRNA携带不同的反密码子，识别不同的密码子，它们称为同工tRNAs，其相对丰度决定了密码子的**选择使用**(codon usage，参阅)。

密码子第三位的简并性可以用Francis Crick的**摆动假说**(wobble hypothesis)来解释，

摆动学说假定在正常碱基配对原则中密码子中第三个碱基和反密码子第一个碱基(**摆动位置**)配对较松散,允许一种 tRNA 识别不同的密码子。这种假说假定正常碱基在摇摆位置时识别能力降低,在某些情况下只识别碱基类型(嘌呤或嘧啶),而不是相对链的特殊的碱基。这解释了两种密码子组成家族的简并性,它们的第三位不是嘌呤就是嘧啶(AAR 编码赖氨酸,AGY 编码丝氨酸)。在另一些情况下,第三位碱基是不相关的,这解释了四个密码子家族简并性的存在(如 CCN 编码脯氨酸)。此外 tRNA 中的稀有碱基允许非特异的相互作用(如次黄嘌呤可以腺嘌呤、胞嘧啶和尿嘧啶配对)。摆动规则在不同物种中有所不同,反映了 tRNA 修饰的差异。

密码子的使用　　**密码子的使用,选择,偏爱和倾向性**(codon usage, choich, bias or preference)指在特殊的生物中密码子家族中的某些密码子被更多使用的现象。这随物种不同而不同,所以简并的氨基酸通常仅由代表它们的密码子中的一种或几种编码,而且在不同生物体中主要的密码子不同。第一和第二位碱基倾向性反映了不同同工 tRNA 种类的相对丰度。第三位碱基倾向性反映了在复杂的摆动规则中某种配对构象更稳定。不管是哪种情况,密码子的使用可以作为基因调控的一种机制(如一个稀有密码子可以延迟蛋白合成)。密码子的偏爱可以是由于总体效应所引起,如嗜热菌偏爱包含鸟嘌呤和胞嘧啶的密码子来保持较高的 GC 含量。

遗传密码的不确定性　　遗传密码在大部分情况下是明确的,这种属性对遗传信息的忠实翻译非常重要。但存在两种允许密码子不确定性的特定环境,因其情况的特殊性,所以并不影响翻译的保真度。

翻译起始中可以观察到不确定性。通用起始密码子 AUG 在内部位点和起始位置编码甲硫氨酸(原核生物和细胞器中起始甲硫氨酸被修饰成 *N*-甲酰甲硫氨酸)。在细菌中,还可以选择性使用其他起始密码子:如在深黄色微球菌(*Micrococcus luteus*)中一般为 GUG,在大肠杆菌中偶尔为 GUG 和 UUG(这些密码子在内部位点为缬氨酸和亮氨酸,但在起始位点总是 *N*-甲酰甲硫氨酸)。在真核生物基因组中很少有不同的起始密码子,虽偶尔用 CUG。这种类型的不确定性反映了蛋白合成中起始和延伸阶段的不同分子环境——在大肠杆菌中甲硫氨酸不插入内部的 GUG 和 UUG 位点(详见蛋白合成一章)。

第二种情况是在由内部插入稀有的硒代半胱氨酸。这种氨基酸与半胱氨酸类似,以硒代替了硫,为有些称为**硒蛋白**(selenoproteins)或**硒酶**(selenoenzymes)的几种基因产物的有效功能所需。大多数蛋白质中不常用氨基酸是翻译后加工的产物,但硒代半胱氨酸是参入前由丝氨酸的修饰产生,然后与其对应的 tRNA 结合。硒代半胱氨酸由通常为终止密码子的 UGA 定义。在编码硒蛋白的 mRNA 中,称为**硒代半胱氨酸插入序列**(selenocysteine insertion sequence, **SECIS**)的二级结构使蛋白合成装置翻译这种密码子,具体的机制目前还不清楚。在细菌中,SECIS 位于 mRNA 编码区,而在真核生物中,SECIS 位于 3′端非翻译区。与大肠杆菌 SECIS 相互作用的蛋白已被发现。

标准遗传密码的歧化　　遗传密码的通用性是很明显的,被研究的生物物种越多,**密码排布**(codon assignment)的差异也就更多地被发现(表 11.2)。大多数衍生发生在线粒体基因组中,可能与它们编码的蛋白种类较少有关。但在植物细胞器中,RNA 编辑(参见)是主要的,目前还不清楚植物中遗传密码的歧化是真的变异还是翻译前 RNA 编辑的结果。

细菌基因组和真核生物细胞核基因组中密码子偶尔也有改变,但常发生在终止密码子处。这些改变的系统分布表明密码子仍在演变。除了这些固定的改变,还有位点专一性的密码子排布的改变,如特殊的密码子因为其位置的缘故,以非常规方式被打断。如上文所讨论的,这样的效应包括在 UGA 密码子处插入硒代半胱氨酸,终止密码子通读,阅读框的移动和绕路(参见翻译调节)。RNA 编辑(参阅)也被认为是正常遗传密码的一种歧化。

表 11.2 密码子排布的变化

密码子	通用的翻译	系统	例外系统中的翻译
AAA	Lys	一些动物的线粒体 果蝇线粒体	Asp
AGR	Arg	大部分动物线粒体 脊椎动物线粒体	Ser
AUA	Ile	一些动物和酵母线粒体	Ser
CUG	Leu	圆柱念珠菌(*Candida cylindracea*)细胞核基因组	终止
CUN	Leu	酵母线粒体	Thr
UAR	终止	一些纤毛虫(ciliate)细胞核基因组[如四膜虫(*Tetrahymena*)]	Glu
UGA	终止	动物和酵母线粒体基因组 支原体基因组	Try
UGA	终止	一些纤毛虫(ciliate)细胞核基因组[如游纤虫属(*Euplotes*)]	Cys

第二种遗传密码 tRNA 对核酸顺序信息的识别是遗传密码的基础。其他生物体系也依赖于对核酸顺序信息的识别,主要是控制转录和其他 DNA 功能的蛋白与 DNA 的结合。对控制序列专一性的蛋白-核酸相互作用的“氨基酸密码”的研究表明不存在通用的序列对序列的对应关系。然而对某些蛋白家族而言,识别的密码已开始被描述(参见核酸结合蛋白)。

(徐 沁 译)

第 12 章
基因组与作图

基本概念和定义

- **基因组**(genome)是细胞中全部的遗传信息,包含了细胞发挥功能所需的“程序”。它被认为是所有的遗传物质,或者当同样信息存在多个拷贝时,单个信息拷贝的遗传物质(后者有时称为**单倍体基因组**)。细胞基因组中多余的拷贝数称为染色体倍数性(ploidy)。在真核生物中,大于 99% 的 DNA 在**核基因组**(nuclear genome)中,但细胞器中也存在 DNA(参见细胞器基因组)。细菌和细胞器的基因组较小,通常是单个环状染色体,但也有报道存在一些线性的细菌基因组。真核生物核基因组相对大得多,分成多个线性染色体。病毒基因组结构变化很大(进一步讨论见病毒一章)。
- 基因组不是简单的基因随机组合。它们有功能性的高级结构,能够根据它们的物理-化学性质和顺序组成得到描绘。大部分生物的 DNA 可以分为几种序列成分:只出现一次的**单一顺序 DNA**(unique sequence DNA)和不同类型的**重复 DNA**(repetitive DNA)。大多数基因是在单一顺序 DNA 中发现,但在中等重复 DNA 中也有与高度保守的多基因家族相对应的基因。其他重复 DNA 顺序不被转录,由分散分布的重复顺序(经常对应于活性的或突变的转座因子)或真核生物中简单顺序的随机重复(其中一些在染色体功能中起作用)组成。
- 细菌和真核生物基因组及高等和低等真核生物基因组之间在基因结构和基因在基因组中的组织方式存在显著不同。细菌和很多低等真核生物有高度复杂和高基因密度的小基因组:基因组的绝大部分是可表达的单一 DNA 顺序。基因小且一般没有内含子。相反,高级真核生物基因组大,主要是非编码的 DNA,可以是单一或重复顺序。基因在大小上变化很大,经常包含多个内含子,内含子一般比外显子大。基因间距离较大。细菌中,基因经常根据功能成簇排列成操纵子,但在真核生物中很少如此。脊椎动物基因组对应于染色体带型,在基因密度上有区域差别。在一些细菌中,基因方向反映了复制起始的位置。
- 目前在国际范围内正对人类和选择的**模式生物**(model organisms)基因组进行作图和测序。这包括脊椎动物,如小鼠和河豚鱼,它们的基因组作图除了本身的用处,其图谱还可用于人类基因组计划的探索(参见比较作图),像大肠杆菌、啤酒酿酒酵母、线虫和黑腹果蝇等物种作为实验室模型广泛用于不同生物系统的研究,是分子生物学研究的基础。有三类重要的基因组作图:细胞遗传、遗传学和物理作图,其分辨率渐次增加。最终物理作图是基因组测序(分辨率是 1 个核苷酸)。一个完整

的基因组序列是无价的,因为它提供了基因结构、调节、功能和表达的信息,不同有机体进化关系,基因组高等结构的性质和基因组进化。基因组序列有很多商业应用,如药物开发、工业生产的疫苗和酶。基因图谱是定位克隆(positional cloing,参阅)的先决条件。在基因密度低的高度真核生物基因组中,集中分析表达的 DNA 十分有用(参见转录作图)。

12.1 基因组、染色体倍数性和染色体数目

染色体倍数性 一个细胞中特定基因的拷贝数目定义为其**剂量**,而整个基因组的拷贝数被定义为细胞的**倍数性**(ploidy)。在真核生物中这就是染色体组的数目。真核细胞如果只有一组染色体则是**单倍体**(haploid)[1],有两组则为**二倍体**(diploid)。在细胞周期中真核细胞的倍数性会发生改变。在 DNA 复制中,染色体倍数加倍,而在减数分裂中减半。增殖细胞的正常染色体倍数是根据它出生时染色体的套数决定。由于嵌套式复制(参见 Helmstetter-Cooper 模型)细菌细胞的有效染色体倍数随生长速度改变,接近复制起点的基因剂量尤其大。

染色体数目 在真核生物中,**一倍体数目**(monoploid number)(x)是代表基因组一份拷贝的染色体的数目,也就是一套染色体中染色体的数目。**单倍数**(haploid number)(n)是配子中染色体的数目。在大多数真核生物中,配子包含一套染色体,也就是 $n=x$,但对于往往是多倍体的植物,n 是多份 x。**二倍体数目**(diploid number)($2n$)通常用来描述大多数动物体细胞中染色体总数,是核型的基础(参见下文)。**C-值**是单倍体基因组中 DNA 的总量,它可以用碱基对、相对分子量或准确分子量来表示。偶尔,染色体倍数性可以以 C-值表示,如二倍体细胞是 2C。

核型(karyotype)是描述总染色体数目和性染色体构象的简单方法。例如,人类男性体细胞的核型为 46,XY,女性是 46,XX。在异常细胞中,核型增加的信息表示了特定染色体异常(参见表 4.1)。**核型图**(karyogram)在另一方面是同源配对染色体染色图谱,用来鉴定染色体异常 (参见染色体分带)。

12.2 基因组的物理化学性质

碱基组成 基因组可以根据它们的物理化学性质来描绘。既然所有细胞基因组都是 DNA,基因组之间的物理化学性质的差异反映了两种性质之一:**碱基组成**(base composition)的大量差异(腺嘌呤/胸腺嘧啶和胞嘧啶/鸟嘌呤的相对量)或是 DNA 甲基化(参阅)

1)单倍体这个术语严格说来是"染色体倍数的一半"且被固定用来描述配子的状态(其染色体倍数是减数分裂细胞的一半)。由于大部分减数分裂细胞都是二倍体,大部分单倍体细胞具有一套染色体,因此这个术语被用于描述一组染色体的细胞。但带有 6 套染色体植物的配子严格说来也应称为单倍体(monoploid),虽然这些配子具有三套染色体,也是三倍体。一倍体术语是特指那些只具有一套染色体的细胞(见第 4 章染色体突变)。

程度的不同。A:T 碱基对包含两个氢键而 G:C 对有三个。G:C 碱基对含量高增加了基因组的物理稳定性，因为它需要更多的能量才能将链分开。高 G:C 碱基对含量与高热熔解温度有关(参见核酸杂交)。同样 G:C 碱基对的相对分子量要大于 A:T 碱基对，GC 丰富的 DNA 的浮力密度大于 AT 丰富的 DNA。甲基化也能增加浮力密度(参见浮力密度梯度离心、卫星 DNA)

碱基组成可以用两种方法表示。**碱基比例**(base ratio)[也称为**不对称比例**(dis-symetry ratio)或 **Chargaff 比例** (Chargaff ratio)]用于微生物 DNA。它定义为(A+T)/(G+C)，以一数字表示。碱基比例大的物种称为 **AT 型**(AT types)，碱基比例小的称为 **GC 型**(GC types)。**GC 百分比含量**(%GC content)适用于所有基因组。它定义为(G+C)/(A+T+C+G)，以百分数表示。嗜热生物经常是 GC 型(高 GC 含量)，因为 G:C 丰富的基因组有助于在高温下维持 DNA 的双链。不同有机体的 GC 含量在表 12.1 中表示。

表 12.1 一些生物的基因组数据

物种	基因组大小	复杂性(%)	GC 含量(%)	基因数目
λ噬菌体	45kbp	>99	48	100
大肠杆菌	4.7Mbp	99	51	4100
啤酒酿酒酵母	13.5 Mbp	90	41	6300
裂殖非洲粟酒酵母	20 Mbp	90		6000
盘基网柄菌	47 Mbp	70	23	7000
秀丽隐杆线虫	100 Mbp	83		14000
黑腹果蝇	165 Mbp	70	39	12000
河豚鱼	400 Mbp	>90	44	70000
担尼鱼	1.9Gbp			70000
非洲爪蟾	2.9 Gbp	54	50	70000
小鼠	3.3 Gbp	58	41	70000
人	3.3 Gbp	64	40	70000
拟南芥	70Mbp	80		25000

复杂性是以单一 DNA 顺序的百分比来表示的，基因数目是指从基因组测序计划中获得。完全的基因组顺序尚未获得的情况中，复杂性是通过现有顺序的资料进行推测来评价的。

碱基比例和 GC 含量是整个基因组中平均值。但在大多数基因组中碱基组成有变化，使一些区域 AT 丰富，另一些 GC 丰富。在较小的基因组中，碱基组成的区域差别通过**变性作图**(denaturation mapping)鉴定：当基因组部分变性后用电子显微镜观察，AT 丰富区域形成泡结构。AT 丰富和 GC 丰富 DNA 的不同化学行为与哺乳动物染色体观察到的带型有关(参见等容线模型)，两种不同荧光染料的不同结合能力可用来将单个染色体分离(参见流式分离)。非特征性碱基组成的染色体区域的存在经常是物种间水平转移的标志(参见密码子选择)；在细菌中，很多**致病性岛**(pathogenicity islands)—水平转移的致

病性基因—就是以这种方式被鉴定的。

12.3 基因组大小和序列组成

基因组大小和复杂性　　单倍体基因组中总 DNA 的量(C-值)预期随物种生物复杂性的增加而增加,因为较复杂的生物需要更多的基因产物。这在广义上是正确的:脊椎动物一般比无脊椎动物有更多的 DNA,后者比真菌多,真菌比细菌多,而细菌比病毒多。

在每个门中最小的基因组的大小看来是根据生物复杂性增加,但在类似的物种中C-值也有反常的差异,在每个门中产生 C-值的扩展。在两栖动物的极端例子中,最小和最大基因组有两个数量级的差别。更有甚者,最大的昆虫基因组比最大的哺乳动物基因组大,而所有物种中最大基因组是显花植物。这样的现象不能仅用所需要的基因产物来解释,集中显示了 **C-值矛盾**(C-value paradox)。

这种矛盾可以用很多真核生物基因组中主要是**非编码 DNA** 来解释。非编码 DNA 可能是重复 DNA 或单一顺序 DNA 中都有发生。基因组的**复杂性**(complexity)由单一顺序 DNA 的总和来定义,可以用物理单位 (参见*碱基对、皮克*)或更经常是总基因组的百分比来表示(表 12.1)。重复 DNA 的存在最早是通过**重新组合动力学**(文框 12.1)被发现并部分解释了 C-值矛盾。同一门类中 C-值的差异主要反映了对基因组复杂性没有贡献的重复顺序 DNA 的含量的差异。当将重复顺序 DNA 考虑在内时,在有类似生物复杂性的物种间仍存在基因组大小的不一致性,特别是在一群单细胞有机体中间进行比较时。例如酿酒酵母中的 C-值大约为 13.5Mbp,而另一种酵母,裂殖酵母接近 20Mbp。这两种有机体有类似的结构复杂性和较少的重复顺序 DNA。差异反映了非编码的单一顺序 DNA 之间的不同,如基因间 DNA 片段和内含子:裂殖酵母 40%的基因有内含子,而酿酒酵母只有 4%基因有内含子。在更高等的真核生物中,基因间区域和内含子更大,内含子数量更多,使基因的平均大小和基因间距离增加。

DNA 序列成分的分布状态和功能　　在细菌中基因组的绝大部分是单一序列 DNA,代表基因和调节元件。一些基因和其他序列是重复的,但**拷贝数(或重复频率)**一般很低,经常<10。这样的重复序列的例子包括 rRNA 基因(在大肠杆菌中有 7 个)和可转座的因子如 IS 因子(参阅)。偶尔特定的序列基序也可被适度重复。在 1.8Mbp 的流感嗜血杆菌(*Haemophilus influenzae*)基因组中,有大约 1500 个拷贝的 30bp DNA 吸收位点(参见*转化*)。类似的大肠杆菌基因组包含了两个重复元件 ERIC(肠细菌重复内在共有顺序)和 REP(重复基因外回文顺序)的多个拷贝。总之细菌重复顺序 DNA<1%,它的基因组大小直接反映了其复杂性。

在真核生物中,重复 DNA 占基因组的很大一部分。所占比例在不同物种中差别很大,在某些真核微生物中低至 5%,在哺乳动物中占 40%～50%,而在一些显花植物中高至 80%。最初复性实验(文框 12.1)表明了真核 DNA 可被分为三种组分,**单一序列 DNA**,**中等重复 DNA** 和 **高度重复 DNA**,这将在下文中详细讨论。基因组 DNA 可以根据他们在细胞中的功能来分类(表 12.2)。

表 12.2 真核生物基因组成分根据含量和功能分类

DNA 类型	定　义
根据含量	
单一顺序(单拷贝,低拷贝,非重复顺序 DNA)	每个基因组中顺序出现一次或很少次。包括大部分基因和内含子,调节顺序和其他未知功能的 DNA。
中等重复顺序 DNA	每个基因组中出现 10～10 000 个拷贝。一般是代表高度保守的多基因家族的分散重复顺序(功能基因和假基因)和转座因子。偶尔成簇排列。
高度重复顺序	每个基因组中出现 10 000～1 000 000 个拷贝的序列。一般作为随机重复顺序被发现,一些超丰度的(弥散的)转座因子也属于这类(如 Alu 元件)。
根据功能	
基因 DNA	基因,即可以表达的 DNA。基因 DNA 可以进一步分为 mDNA(编码蛋白)、rDNA、tDNA、snDNA 等,代表了不同的基因产物。
调节 DNA	DNA 的功能是调节基因表达 (如启动子、增强子)或调节 DNA 功能(如复制起始区,核基质结合区域)。
基因内 DNA,间隔 DNA	内含子和分隔基因的 DNA。
卫星 DNA	靠近着丝粒、端粒和其他位置的高度重复 DNA,有些卫星 DNA 在染色体功能中发挥作用。
自在 DNA	功能是介导自身在基因组中的复制和生存,如一些卫星 DNA 和转座因子。
无用 DNA	没有确定功能的 DNA。

12.4　基因结构和高级基因组组织

DNA 序列成分中基因分布　　在高等真核生物中,大多数基因被认为处于单一顺序 DNA 中,因为它们的存在可以通过单一突变被定义。如果基因是重复的,观察到表型突变需要几个突变同时发生,这样会缺乏维持多拷贝的选择压力(参见*功能过剩*)。细菌和低等真核生物的基因致密基因组主要是单一顺序 DNA 组成。追踪重新组合实验首先表明高等真核生物的大多数基因确实在单一 DNA 顺序中,这还被单个克隆基因的 DNA 基因组杂交所证实。

通过逐渐降低杂交分析的严格性,越来越多的基因被发现是中等重复顺序成分,表明基因属于不同程度序列保守的多基因家族。在高严格的条件下,重复顺序 DNA 中被鉴定的基因是高度保守或是完全一样的多基因家族(如 rRNA 基因和组蛋白基因)。属于多基因家族的基因所占比例在不同物种间有差异。在大肠杆菌中,最近发表的基因组序列数据表明在这种细菌中高达一半的基因属于多基因家族,而其他细菌中比例要小得多(如在流感嗜血杆菌中是 25%～30%)。这可能反映了大肠杆菌的不同生活方式,如它们所需的营养条件。在哺乳动物中,真正单一基因的比例还要小,大部分人类基因属于多基因家族,有些家族包含了上千个成员。这可能反映了发生在脊椎动物谱系中的基因组复制和家族扩展(参见第 22 章蛋白质:结构、功能与进化)。基因组中多基因家族的结构和组织在下文中讨论。

基因大小和内含子-外显子结构　　细菌基因较小(平均 1kbp),且在大小上变化不大,而

高等真核生物基因较大(平均 16kbp),且大小变化很大。哺乳动物中最小的基因与细菌基因相当(如人类的 α-干扰素基因<1kbp),但很多超过 100kbp DNA 大小,目前发现的最大基因是人类的肌营养不良蛋白基因,有 2500kbp 长。

尽管高等真核生物基因一般比细菌基因大得多,从它们中得到的 mRNA 并非如此。这样的差异是由内含子(introns)引起的:内含子是打断转录单位的间插序列,必须在 RNA 水平上去除(参见第 27 章 RNA 加工)。转录单位余下部分通过剪接结合起来并表达称为外显子(exons)。基因大小与基因中外显子比例成反比。细菌基因一般缺乏内含子,100%是外显子。内含子在很多真核微生物中也很少(如酿酒酵母),它们的基因一般大小为 1~2kbp 与细菌类似[1]。在人类中,最小的基因有很少的内含子且内含子很小(如 500bp 组蛋白 H4 基因没有内含子)。相反,最大的基因有>95%的内含子。上面提到的肌营养不良蛋白基因有 78 个内含子,平均大小是 30kbp;只有基因的 0.5%是外显子。

然而高等真核生物的内含子大小和数量变化很大,外显子大小则在一狭窄的范围内。如人类的不连续基因中,外显子平均长度为 170bp,在 50bp 和 300bp 范围内变化。有一些值得一提的例外,如载脂蛋白 B 基因的 26 号外显子有 7.6kb 长度,但这样的例子很少。无脊椎动物与脊椎动物相比有更大的外显子,因为它们的内含子极少(表 12.3)。对内含子的起源和功能的讨论,参见第 22 章蛋白质;剪接机制参见第 27 章 RNA 加工。

表 12.3 不同真核生物中内含子-外显子的组织。酿酒酵母很少有被打断的基因,基因长度与 mRNA 长度一致。高度真核生物基因大小逐步加大,但 mRNA 大小保持恒定。一般基因大小与内含子数量成比例,与外显子含量成反比。

物　种	基因平均长度(kbp)	平均内含子/基因	平均 mRNA 长度	%外显子
酿酒酵母	1.5	>95%不被打断	1.5	100
线虫	4	3~4	3	77
果蝇	11	3~4	3	25
人类	16	6~7	2.5	13

注:线虫与果蝇有类似的内含子数量,但内含子更小,使基因平均大小要小一些。

基因数目和密度　几种微生物的基因数目通过基因组测序计划被确定(参见表 12.1 中的例子)。细菌基因数目变化有一个数量级,从 473(枝原体)到大约 8000 (粘液球菌)。而大肠杆菌大约有 4100 个基因。在基因数目上,最大的细菌基因组与低等真核生物相差不多。酿酒酵母有 6340 个基因。无脊椎动物的模式生物果蝇和线虫预计有酿酒酵母两倍的基因数量。脊椎动物代表了另一层次的复杂性,预计有大约 70 000 个基因。这些预计的信息来自部分基因组测序、CpG 岛分析(参阅)和 EST 方案(参阅)。

维持一个独立生命有机体所需的最小基因数目是多少?细菌基因组比较发现了一系列必需生化途径,并有 256 个基因编码其中的成分。真核生物细胞建立复杂的细胞内结

1) 细菌基因经常被定义为开放阅读框架,而真核基因经常被称为转录单位(见第 8 章基因)。酵母基因平均大小比细菌基因大,部分是因为酵母 mRNA 包括了非编码区。

构似乎需要更多的途径。多细胞生物调节发育和分化细胞的功能就需要更多。然而关键生化途径的数目在所有后生动物中类似,因为脊椎动物中大量基因被认为是通过整个基因组的两轮重复,加上不同染色体区域和单个基因的重复产生的。起初过剩的基因被用于特殊的功能,经常是因为表达模式的分化,但途径是高度保守的。有很多关于这种假说的证据趣谈:果蝇中很多单基因出现在脊椎动物中的多基因家族中。例子包括果蝇 *hedgehog*(脊椎动物的 *sonic hedgehog*,*desert hedgehog* 和 *indian hedgehog*),果蝇的同源异型复合物(脊椎动物的 *Hox-A*、*Hox-B*、*Hox-C*、*Hox-D* 基因簇),以及编码信号蛋白 Ras 和 Raf 的基因。

基因组中基因的高度有序组织可以分为结构组织和功能组织、方向和基因密度。枝原体基因组中的一个独特的性质是所有基因的方向都是与复制方向相反;这种现象的意义目前还不清楚。细菌和真核生物基因组最大的不同是细菌中功能相关而结构无关的基因组织成操纵子。这类组织在真核生物中明显缺乏,因为多开放阅读框架不能从一共同的多顺反子转录本中翻译得到(参见操纵子、多顺反子 mRNA、核糖体结合位点)。代替它的是,功能相关的真核基因,如那些相同生化途径的成分的编码基因通常是分散的。因为串联重复,在细菌和真核生物中都有结构相关基因的成簇排列,这样的基因可能是功能相关的和处于相关的转录调控下(如珠蛋白基因簇和 *Hox* 基因簇)。

细菌中**基因密度**最高,平均每一 kbp DNA 有一个基因,反映了操纵子中开放阅读框架排列紧密。酿酒酵母中的基因密度大约是细菌中的一半,反映了转录单位的分离。高等真核生物中基因密度的降低有几个原因。第一,在越复杂的有机体中,内含子越大也越多。第二,重复序列 DNA 所占比例也逐渐增加。第三,因为需要更大和更复杂的调节元件,基因间距离增加。在这些一般原则中也有例外,例如河豚鱼(*Fugu rubripes*)。这种东方豚鱼的基因组值得注意的特点是高复杂性(>90%)和基因小,反映了缺乏分散的重复顺序和基因组范围内内含子大小的减小。东方豚鱼中与哺乳动物大基因的同源基因,如肌营养不良基因只有约 10% 的大小,但保留相同的内含子-外显子结构;内含子小得多,大多数在 50bp 和 150bp 之间。

基因密度在整个高等真核生物中并不是不变的。哺乳动物基因组基因密度有区域性变化,反映了等容线结构(参见等体容线模型)。无脊椎动物基因组结构较简单:果蝇基因组分为 115Mbp 的基因富集区和 50Mbp 的基因缺乏的异染色质区(参见重叠基因、嵌套基因)。

12.5 重复顺序 DNA

多基因家族　　**多基因家族**是一个基因组中同源基因的家族(参见基因家族)。在这些基因中,拷贝数、顺序保守性、构成、分布状态和功能相关性有很大差异。在一些例子中,家族成员可能非常相似或完全一样(如 rRNA 基因),在严格的杂交条件下作为重复 DNA 被鉴定。在其他一些例子中,保守性非常差,即使通过序列比较也不能发现(参见同源性、基因超家族)。经典的多基因家族是结构相似,在整个编码顺序中保守。它们可以在特殊座位上成簇排列(如人类 β-珠蛋白基因)、分散的(如人类肌动蛋白基因)或者两者都有(玉米醇溶蛋白基因)。成簇的多基因家族的偶尔分散的成员称为孤独基因

(orphon)[1]。其他多基因家族只在特殊的对应保守的蛋白结构域的亚基因区域相同(如同源异形基因在编码 DNA 结合的结构域的 180bp 同源盒相关)。更有其他一些只在一个非常短的氨基酸基序相关(如 MADS 盒和 DEAD 盒 RNA 螺旋酶基序)。更为复杂的是,很多基因呈现为对应不同蛋白结构域的相对独立功能单位的嵌合分子,使它们能够同时成为几个不同家族的成员。这样的基因被认为是通过祖先基因间的重组产生的(参见外显子改组),可以包含重复的编码信息(参见外显子重复)。

假基因　多基因家族经常包含结构保守的基因,它们是通过积累突变产生,来满足不同的功能需要。在一些例子中,突变使基因功能完全丧失,这样的无功能的基因拷贝称为**假基因**(pseudogene),经常用希腊字母 ψ 表示。根据起源和结构的不同,假基因分为两类。

未加工的假基因(nonprocess pseudogenes)也称为**常规假基因**(conventional pseudogenes),是通过基因组 DNA 复制产生(参见第 15 章突变与选择),经常位于相同基因有功能拷贝的附近。它们与有功能的同源基因有类似的结构,可以包括内含子和调节元件。这样的假基因在细菌和真核生物中都有发现,因为它们是积累突变,包括使转录消失的调节突变和产生截短编码产物的无义突变,所以能够被识别。偶尔未加工的假基因可以通过一个有利的突变重新激活。产生未加工假基因的过程也可能产生部分基因或截短的拷贝。

加工的假基因(processed pseudogenes)也称为**反转录假基因**(retropseudogenes),是通过对 mRNA 的反转录和获得的 cDNA 的随机整合而产生;它们经常是分散的。加工假基因是由反转录因子(参阅)编码的反转录酶和整合酶的外来活性而产生的,只在真核生物中被发现。加工的假基因结构对应于起源基因的转录单位,缺乏内含子和侧翼顺序。因为缺乏侧翼顺序,加工假基因一般不表达,尽管它们偶尔整合在内源性启动子附近,并受它的控制(人类编码丙酮酸氢化酶的基因被认为是这种方式产生的)。RNA 聚合酶 III 有内在的启动子,所以它的加工的假基因可以表达。人类高度重复 Alu 元件是表达的 RNA 聚合酶 III 加工的假基因的例子。

结构和功能的冗余性　**冗余**(redurndant)序列是在基因组中出现超过一次的序列,也就是增加基因组大小,并不增加复杂性的序列。冗余基因并不必定是功能冗余。一些基因被发现有冗余拷贝,以产生足够基因产物(rRNA 基因属于这一类),另一些进化以实现不同功能。功能冗余可以通过当特定基因或元件缺失造成表型缺失来建立。完全或部分功能基因冗余在多细胞有机体的很多定向突变中可以看到,即使同样的基因在异位表达时表现出显著的功能效应的增加。另一个例子是转录因子 MyoD,它可以通过激活生肌途径使很多不同的细胞类型转变成肌肉。当小鼠 *myoD* 基因删除(参见基因敲除),同源基因无效的个体是正常的。这是因为另一个转录因子,Myf-5 能够对MyoD缺失进行补偿。功能冗余经常反映了结构冗余(祖先基因通过复制产生的两份拷贝,如同上述例

1)孤独基因与**孤儿基因**(orphan)不同,孤儿基因是在基因组测序计划中发现的,在其他有机体中没有对应的基因,已确定它没有功能。

子[1]),它们可以补偿相互的功能缺失)。在其他的情况中,不同的基因在相同的功能上汇集,例如几种不相关的蛋白从两栖动物织原中分泌——腱蛋白,成头蛋白,囊泡抑制素;它们的共同功能是阻断 TGF-β 信号。功能冗余在有重要发育作用的基因中是普遍存在的,而在看家基因中较少。为什么这么多基因功能冗余,它们是如何通过选择保留下来?有限的基因敲除动物研究表明敲除突变在不同的品系(也就是不同的遗传背景)中经常有不同的效果。迄今为止所报道的很多冗余现象反映了不同特殊遗传背景和环境的影响,而自然选择在一不受控制的体系中进行,其中适应性的精细差别可以被应用。

重复 DNA 顺序的结构 重复 DNA 顺序由特定大小序列(**重复单位**,repeat unit),以特定拷贝数目在空间上以特殊的方式组成。重复单位可以以三种方式被组织:**串联重复**(tandem repeats)在单个重复单位间没有间隔;**不完善的重复**(hyphenated repeats)被小间隔分离,但还是成群排列;**分散重复**(dispersed repeats)散布在整个基因组中。单个重复顺序间可以是相同方向(**正向重复**)或者是相反方向(**反向重复**)排列(图 12.1)。

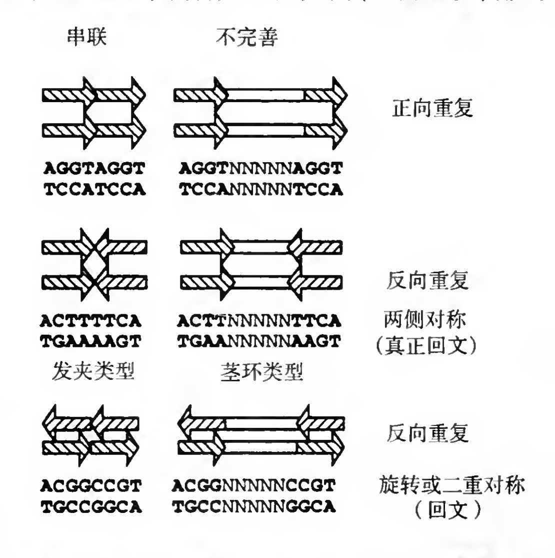


图 12.1 重复 DNA 的顺序结构。重复顺序可以相同方向排列(**正向重复**)或以相反方向排列(**反向重复**)。反向重复中可以是每条链以相对的方向阅读的**两侧对称**(bilateral symmetry)(真正的回文),或者是两条链以相同极性阅读的顺序相同的**二重对称**(dyad symmetry)(也称为回文,尽管有时并不精确)。正向和反向重复都可以是连续的或被非特定顺序所间隔。

串联重复 DNA 中最简单的结构是重复单位只有一个核苷酸,这被称为**同聚体**(homopolymer)。也有二核苷酸,三核苷酸等串联重复,与大重复单位的串联重复一样称为小卫星 DNA(参阅)。分散的单核苷酸或二、三核苷酸的重复显然不值得注意;只有比随机出现频繁得多的长 DNA 顺序的分散重复才有意义。短的分散重复经常对应于功能 DNA 基序(如流感嗜血杆菌上升位点的 30bp),然而转座因子和冗余基因要大得多。

---

1)事实上 MyoD 不能完全补偿 Myf-5 的缺失,因为 *myf*-5 基因在小鼠发育早期表达,功能缺乏延迟肌肉生成直到 *myoD* 表达开始后恢复,结果是推迟了肌肉发育和产期致死。很多发育基因显示出这一类的部分冗余,这些基因具有重叠的功能,当在重组中缺失后比缺失单个的基因产生更严重的表型。在轴骨和肢骨发育中,脊椎动物的 *Hox* 基因是一个有说服力的例子(见文框 6.9),如同 *myoD* 和 *myf*-5 双剔除时产生没有肌肉等组织的小鼠。

短的功能性基序可以由它们特殊的结构鉴定。很多转录因子和其他 DNA 结合蛋白的识别位点由一对正向或反向重复组成，可以是不完善的或串联的，反映了蛋白的二聚性。反向重复经常是二重对称，有形成如发夹和茎环(参阅)二级结构的潜在能力。

作为分散重复 DNA 的转座因子 如上文所讨论的，一些基因组范围分散的重复 DNA 对应于多基因家族的成员，包含功能基因和假基因。另外它可以代表在 DNA 水平上起作用的基序。大多数分散重复 DNA 对应于有功能的转座因子或它们的"空壳"形式(通过突变失活的因子)。这种序列类型的优势在不同生物体中变化很大。在细菌基因组中，转座因子的拷贝数经常<10，而脊椎动物一般分布广泛(尽管在河豚鱼基因组中不存在)。在哺乳动物中，两类特殊的逆转录因子是如此丰富，以致它们每个都定义了一类不同类型的分散重复。**SINEs 是短散布核元件**(short interspersed nuclear element)，对应于加工的 7SL RNA 假基因的拷贝，它在人类中称为 ***Alu* 元件**，在小鼠中是 B1 元件。**LINEs 是长散布核元件**(long interspersed nuclear elements)，对应于称为 LINE－1(L1)的丰富反转录转座子的拷贝。*Alu* 元件大约 300bp 长度，像其他转座因子一样两侧是正向重复，反映了其整合机制(参见移动遗传因子)。它主要位于 GC 丰富的 DNA 区域，估计有 10^6 的拷贝数，平均每 4kbp 有一个元件。相反 L1 元件最大长度为 6kbp，拷贝数 10^5(尽管全长的元件只占一小部分，<5000 拷贝)。L1 和 *Alu* 元件都是与基因相联系的，但它们的分布是相关的，与基因组的等容线结构有关(参阅)，可能是因为整合靶位点的倾向性。这些元件都不在基因的编码区——它们经常出现在内含子和两侧区域，*Alu* 元件偶尔出现在基因的 3′非翻译区域，可以被 RNA 聚合酶 II 作为基因的一部分转录。

卫星 DNA 是高等真核生物基因组重复程度最高的成分，通过快速再聚集鉴定，由非常短的串联多次重复 DNA 序列组成。占优势的高度重复 DNA 在物种间变化，但一般占了基因组的 10%～30%。因为它的低复杂性，有时称为**简单序列 DNA**，又因为其不寻常的核苷酸组成，它经常在浮力密度梯度离心中从整个基因组 DNA 中分离成一个或多个"卫星"条带，也称为卫星 DNA。**隐蔽卫星 DNA**(cryptic satellite DNA)有与大多数基因组 DNA 相当的浮力密度，不形成卫星条带；它通过其他方法被鉴定，如限制性作图。卫星 DNA 以大的基因簇(100～3000kb)分布，经常位于异染色质的着丝粒，可能在染色体功能中起作用。大多数人类染色体的中心粒 DNA 包含了隐蔽卫星 DNA，称为**阿尔法 DNA (α－卫星 DNA)**，尽管另一种成分，**β－卫星 DNA** 至少在人类 8 条染色体的中心粒中也很丰富。α－和 β－卫星 DNA 家族中染色体特异性序列存在差异。

在昆虫中，卫星 DNA 由很多非常短的显著链不对称序列(5～15bp)组成。哺乳动物卫星 DNA 的组织方式更复杂。简单重复序列表现出一些可变性，经常形成些串联重复的一定程度可变的区域。卫星 DNA 因此是由分层结构的简单序列块组成，被认为通过持续突变和扩增的循环产生，可能涉及不对称交换和基因转换(见第 15 章突变和选择)。

小卫星 DNA 和微卫星 DNA 大多数卫星 DNA 是以染色体着丝粒区域或核仁组织者(参阅)的重复序列组成的大基因簇存在，但也经常出现在称为**小卫星 DNA** 的小基因簇(100bp～10kbp)中，一般位于端粒处。有两种形式的小卫星 DNA。在每个染色体臂的末端是端粒 DNA。在大多数真核生物中，它由特征性的几千碱基的串联五核苷酸或六核苷酸 DNA 重复组成(见表 5.5)，它的功能是在随后的 DNA 复制周期中阻止染色体缺损(参

见端粒、端粒酶)。第二类**高度可变的小卫星 DNA** 位于亚端粒区域。高度可变 DNA 的重复单位在不同的位置不同,但都包含了共同的 GC 丰富的核心共有序列。每个位置的拷贝数是高度多态性的[因此又称为 **VNTR 序列**(variable number of tandem repeats,同向重复序列可变数]。高度可变小卫星 DNA 的功能(VNTR DNA)还不清楚,但它可能可以促进重组(在染色体的亚端粒区域交换趋向于成簇)。端粒位置的倾向性意味着小卫星 DNA 不仅对基因组范围的遗传作图有用,它还可被广泛用于 DNA 印记的诊断标记(文框 12.2)。

微卫星 DNA 出现在更小的基因簇(<200bp)中,以非常短的重复单位(1~4bp)为特征。它们有高度的多态性,分布在整个基因组中,所以它们是理想的遗传标记。在两种可能的同源多聚体中,ploy(A)/poly(T)远比 ploy(C)/poly(G)普遍,且二核苷酸微卫星 ploy(CG)/poly(GC)因为 CpG 基序的损耗(参见 5-甲基胞嘧啶)而稀少。三和四核苷酸微卫星 DNA 相对稀少,但作为标记比通常出现的二核苷酸微卫星更有用,因为在 PCR 基因型印记中链的跳格较少(参见分子标记、链跳格)。

一般小卫星和微卫星 DNA 是在基因外的,但偶尔出现在基因的编码区中,产生高度多态性的蛋白;在有些情况中这是病理性的(参见三联体重复综合征)。其他重复 DNA 更稳定的形式也可以在基因中看到。α2 胶原基因包含很多由 9bp 单位重复组成的外显子,产生每三个残基有一个甘氨酸的特征氨基酸顺序。大的重复单位也被发现,最大的对应于整个蛋白质结构域(参阅)。

12.6 哺乳动物基因组的等容线组织

染色质的双相结构 对哺乳动物染色体转录活性和染色质结构的研究表明染色质在有潜在或真实的转录活性时与抑制状态的染色质采用了不同的结构,反映了核苷酸修饰和组织方式的差异和高级折叠结构的去除(见第 3 章染色质)。这一现象的两种结果是活性染色质在 S 期早期复制,变得对 DNase I 敏感。通过早期复制 DNA(复制分带)细胞遗传作图或 DNase I 敏感区域(D 带型),对激活和抑制染色质结构域分布分析,发现在观察到的图谱与通过常规染色体分带技术(参阅) 得到的带型,如 G-带,有惊人的相关性。这样的技术被认为是显示了不同碱基组成 DNA 区域的区别,暗示了哺乳动物基因组物理化学性质和功能性质的相关性。进一步的研究确实显示了 GC 丰富、G 带染色淡的区域基因丰富,而 AT 丰富、G 带染色深的区域基因相对少。在人类基因组中,*Alu* 元件可能也倾向于位于浅带区域,而 L1 元件倾向于深 G 带区域。

等容线和高级基因组结构 等容线模型(isochore model)将哺乳动物基因组分成不同的区域,长度>300kbp,以相对均一的碱基组成为特征。哺乳动物基因组平均的 GC 含量约 40%,但不同区域间在 37%和 55%之间变化。**片段化** DNA 可以通过浮力密度梯度分为**五个等容线类型**:L1 和 L2(AT 丰富)和 H1、H2、H3(GC 丰富)。所有的哺乳动物显示了类似的等容线表现。

通过确定克隆基因的 GC 含量和将 YACs 分成等容线类型,可以研究等容线类型中基因分布。AT 丰富的等容线组成了人类基因组的 65%,但只包含 30%的基因。在 H2 和 H3 等容线中基因密度最大。在 H3 等容线类型中,预计密度是每 10kbp 一个基因,与

果蝇基因组的平均基因密度相同,但比酿酒酵母基因组少五倍。相反,在L2等容线类型中,预计是每100kbp一个基因。令人吃惊的是,从基因组计划数据有限的分析中显示GC丰富的等容线相对小的基因丰富,一般有少而小的内含子和较小的开放阅读框架,而AT丰富的等容线倾向于包含有很多大内含子的较大基因。

杂交分析确定了等容线的分布反映了哺乳动物染色体的带型,AT丰富的等容线大致对应于深色G带而GC丰富等容线对应于淡G带。基因密度非常大的H3等容线倾向于在亚端粒区域,尽管也有例外:在人类基因组中,19号染色体几乎整个都是由H3等容线组成,而13号染色体几乎都是L1和L2等容线。*Alu*和LINE元件差异分布的现象被证实:*Alu*元件倾向于出现在GC丰富等容线中或邻近位置,而LINE元件与AT丰富等容线有联系。有趣的是*Alu*和LINE元件可能在等容线边缘起结构的作用,因为它们经常在这些位置的邻近基因簇中被发现。

12.7 基因作图

基因作图的目的　　**基因作图**是将基因分配到染色体的基因座上。通过对很多基因和其他标记相对位置的作图,可能产生一个**染色体图谱**或整个基因组的图谱。基因作图可帮助人们理解与开发生物性状的遗传性质,特别是通过遗传连锁(参阅)将一种性状的遗传与另一种(或者合适的遗传标记,参阅)相联系,或者通过将表型差异与染色体结构的改变联系起来。商业上,这可以使动物和植物繁殖产生有改良性状的农作物和兽类,在人类范围中,这使生化机制不清楚的遗传疾病的基因在特定染色体或染色体带型中**定位**(分配到一个基因座),并通过标记相连锁追踪家谱(参见基因追踪)。

根据最近重组DNA技术的进展(参见第24章重组DNA、第21章聚合酶链式反应),通过建立基因组DNA片段克隆和分析可能产生详细的基因图谱。最终,可能获得基因组的整个DNA顺序,这一无价的资源不仅提供了基因结构的信息,而且有高度有序的基因组序列高级结构和基因与基因组进化的信息。现在,对几种模式生物在国际范围内开展制作遗传和物理图谱的工作,最终的目标是确定整个基因组的顺序(表12.4)。与物理基因作图相联系的技术用于定位克隆(参阅),每个基因序列的获得有很多商业用途,包括药物开发。基因组作图、分析和测序的学科称为**基因组学**(genomics)。

三种基因图　　有三种主要的基因图:遗传(连锁)图、细胞遗传图和物理(分子)图。

遗传或**连锁作图**(genetic or linkage map)是将基因分配到**连锁群**(linkage group)(因在染色体上位置邻近倾向于一起遗传的一群基因)。遗传图以绝对单位被标定,反映了通过重组(如在减数分裂作图中)或染色体片段化(如在放射性杂交作图和HAPPY作图中)标记分离的可能性,遗传作图依赖的原理是两个基因座出现在同一DNA片段的机会随它们之间距离的增加而减少。在细菌中的遗传作图涉及相似的原理,水平基因转移自然机制使作图成为可能(见第10章细菌中的基因转移)。遗传图谱分辨率在细胞遗传和物理图谱之间,但那些基于重组频率的作图可以被相对高或相对低的重组所歪曲(参见重组热点和冷点)。另外,重组的频率在各种物种中并不一样;因此遗传图谱在不同的物种中反映了不同的物理距离。

细胞遗传图谱是通过将可观察到的染色体重排和表型相联系而产生的。这一类图谱

分辨率低，在少数物种中应用，如哺乳动物和果蝇，它们在自然或诱导可以产生可重复的染色体带型(另见病理作图)。简单基因组的细胞遗传作图(如病毒、质粒等)用来确定它们物理性质差异的位置(参见变性作图)。

表 12.4 用于基因组作图和测序计划的物种

生物体	评论
细菌	大肠杆菌和枯草杆菌是得到细致研究的模式生物，被用于基因结构和调节的前沿研究。大肠杆菌基因组序列最近被发表，而枯草杆菌在近些年也将完成。许多其他细菌基因组已经完成测序：第一个是嗜血流感菌。
酿酒酵母	酿酒酵母和相关的裂殖酵母广泛用于真核基因结构和调节的研究，在核心途径如 DNA 复制和细胞周期的研究中尤其重要。最近完整的酿酒酵母序列被发表——这种酵母有大约 6300 个基因；比原先预测的多。
线虫	线虫是最简单的无脊椎动物模式动物(基因组为 100Mb)。它被广泛用于发育、特别是神经发育的研究，这是因为它不变的世系(从一个卵细胞成为 959 个体细胞的成体的发育途径已被完全阐述)，且其神经系统的相互作用被完全作图。已经完成高密度的遗传和物理作图，测序计划也接近完成。
果蝇	果蝇是所有模式生物中研究得最广泛的，特别是在发育遗传的早期研究中。基因组用形态标记高密度作图，果蝇还被广泛用于连锁分析原理的教学。基因组大小为 165Mb，测序计划处于起始阶段。
东方豚鱼	河豚鱼与大多数脊椎动物相比基因组较小(400Mb)，但基因数目相当，也就是大小的降低反映了非编码 DNA 的减少(特别是分散的重复 DNA 和内含子)。基因和调节序列表现出高度保守，所以河豚鱼基因组是鉴定其他基因密度低的脊椎动物基因的潜在有力工具。
斑马鱼	斑马鱼基因组是典型哺乳动物基因组的三分之二大小，正在进行遗传和物理作图。斑马鱼因其遗传可操作性，是有用的发育模式，如单倍体胚胎可存活到孵育期。最近斑马鱼用于饱和突变筛选发育相关基因
小鼠	有许多原因，小鼠是直接与人类相关的最重要的模式生物(或者较小范围中大鼠也是)。小鼠和人类基因组在大小、复杂性可以相比，大量的同线性是保守的。小鼠的遗传可操作性在哺乳动物中是无可比拟的，因为它个体小，容易繁殖。它也可以进行突变筛选和遗传杂交产生标记图谱和对疾病基因座定位。它还可以对生殖细胞修饰来研究基因功能(参见转基因动物，基因敲除，ES 细胞)
其他脊椎动物	对许多其他脊椎动物也进行了高密度标记作图和基因作图，包括很多重要商业价值的家畜(牛、猪、羊)和驯服的动物(猫、狗)。通过用标记辅助对质量性状的筛选，来改进家畜的产量和质量，并作为比较基因组学的数据(参阅)。
人类	**人类基因组计划**(Human Genome Project)希望在 2005 年前完成，目标是完成对基因组的遗传和物理作图，确定整个基因组序列。这些信息的提供可以给医药、诊断和开发药物带来希望，而不仅仅是确定大多数遗传失常和疾病的机制。但也存在这些信息被错误使用的担心，如对有特殊基因型个体的歧视，可能导致种族灭绝的重新出现。对个别基因的作图通过表达顺序标记作图(参阅)正在进行中(在基因组中的 3%)。
拟南芥	拟南芥在植物中有相对紧密的基因组，它与果蝇位于同一类，也就是相对简单的后生模式生物。与果蝇一样，这种植物被广泛用于发育的研究，遗传和物理作图正在构建中。
其他植物	有几种重要商业价值的植物物种基因组计划正在进行，如稻、小麦和玉米，这将有助于选择改良的性状，如抗虫、抗除草剂等，及增加产量。这些性状中很多是数量性状，标记辅助筛选将会帮助在繁殖过程中发现这样的性状。

物理或分子图谱是通过基因组 DNA 克隆片段的排序产生的，以真实单位(碱基对、千碱基对、兆碱基对)标定。物理作图有最精细的分辨率，是作图计划的最终目的。然而由于高等真核生物基因组只有一小部分被表达，有些物理作图方法被用来鉴定转录的序列(也就是基因)。

在过去，精细的策略被用来产生不同生物的连锁图谱，但这一方法逐步被粗放的物理作图方法所代替，因为物理方法获得整个基因组顺序相对容易。遗传作图在细菌和单细胞真核生物中已经落伍，因为物理作图和完全基因组测序相对快得多。在不久的将来，遗传作图在如线虫和果蝇等物种(这些物种最先用于遗传作图)中也会变得陈旧，但在大基因组的起始的低分辨率作图中还是需要的，如人类基因组。分子标记(参阅)的改进使三类作图逐步综合。多态性分子标记，如限制性片段长度多态性和包含小卫星 DNA 序列标记位点可以被同时用于连锁分析中的遗传标记和作为核酸探针鉴定物理克隆。同时，原位杂交使物理克隆能精确分配到细胞遗传作图的染色体条带上。

12.8 遗传作图

用自然和人工断点遗传作图　遗传作图的基本原理是，假定染色体断裂是随机的染色体上两个任意基因座越远，它们越容易被染色体断裂所分离。

利用这原理通过人工导入染色体断裂点发展了几种作图技术，包括**放射性杂交作图**(radiation hybrid mapping，染色体断点是通过放射体细胞杂交产生)和 **HAPPY 作图**(基因组 DNA 是通过振荡或超声打断)。在每一种情况中，物理作图技术可以被用于检测连锁：同一 DNA 片段上是否存在两个标记是通过杂交、PCR 或基因产物的检测确定的(图 12.2)。实验被多次重复，连锁率通过标记被分离的频率计算。放射杂交作图是在射线百分数中(cR_x，x 是 X 射线的剂量)校正，1 cR 相当于 1%频率的分离。

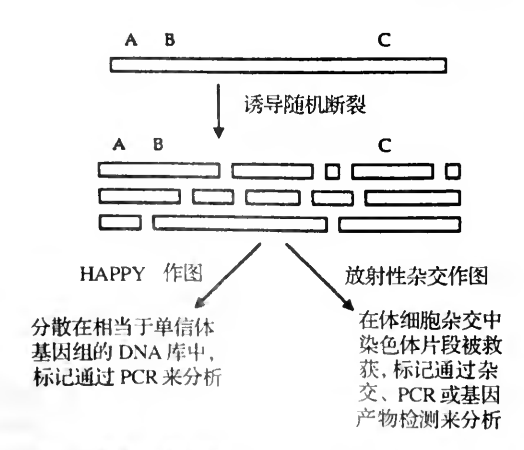


图 12.2　用人工导入的断裂点遗传连锁作图　在放射性杂交作图中染色体的断裂是通过照射诱导产生的，而且在体细胞杂交(**放射杂种**)染色体片段被获救。它们可以通过杂交、PCR 或蛋白质检测来分析。在 HAPPY 作图中(单倍体 PCR 作图)，基因组 DNA 被随机剪切分散在相当于单倍体基因组的 DNA 库中，标记通过 PCR 来分析。

连锁作图最通常的方法是利用因为减数分裂时交换(同源染色体)产生的天然染色体

断点。在这种情况下，物理距离是根据标记间重组频率标定的(如距离越远，它们之间越容易发生交换)，这要求在两个基因座上是杂合基因型，这样亲本的和重组的单倍型可以被区分。在图 12.3 中举了个例子。与减数分裂连锁分析相关有很多限制和问题，包括样本大小、基因型检测和所提供标记的数量，信息性分离的缺乏(在人类中)和内在的不精确性(文框 12.3)。

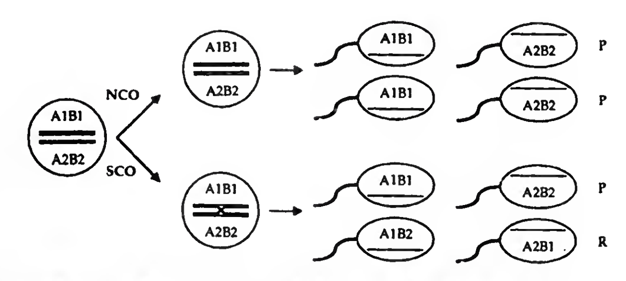


图 12.3 用重组频率进行遗传作图的原理：考虑两个连锁的基因座，A 和 B。一个个体(图左)在两个基因座上都是杂合的(如基因型为 A1A2B1B2)，可以从一个亲本遗传得到 A1 - B1 单元型，从另一个亲本得到 A2 - B2 单元型——这些是**亲代(非重组)单元型**，也就是减数分裂输入。进行减数分裂的个体转移一条染色体到他的后代；减数分裂的四种产物在图中表示。如果 A 和 B 之间没有交换(NCO)，作为减数分裂的输出，后代继承了亲代单元型的一种(A1 - B1 或 A2 - B2)，称**为亲组合的或非重组合**(P)。相反的，如果 A 和 B 存在单个交换(SCO)，作为减数分裂的输出，后代继承了重组的单倍型(A1 - B2 或 A2 - B1)，称为**重组合**(R)。随着基因座之间距离的增加重组体基因型的频率增加，因为交换在距离大的基因座之间更容易发生。然而，重组频率最大是 50%，因为四条染色体中只有两条涉及单交换，所以只有 50% 的减数分裂的产物是重组的(文框 12.3)。注意如果个体在两个基因座上都是纯合的，减数分裂将**不提供重组信息**，因为不可能区分亲组合和重组合。在二倍体物种中，成功的减数作图依赖于根据后代表型确定单元型的能力。如果亲代提供同样的等位基因或者等位基因间是显性关系，这也是不可能的(见正文)。

连锁分析中的样本大小　如果从给定交配得到足够的后代可以被**计算**和**定型**(确定它们的基因型)，任何一对杂合基因座之间的**重组频率**(r)(recombination)或者**交换值**(cross-over value)可以用以下公式计算：

$$r = \frac{\text{减数分裂的重组产物}}{\text{减数分裂的总产物}}$$

重组频率被用来标定遗传图谱：1% 的重组频率(也就是每 100 个后代只有一个遗传了重组单元型)对应于一个**遗传图距单位**或**厘摩**(根据 Thomas Hunt Morgan 命名，在他的实验室最先进行果蝇基因的遗传作图)的遗传距离，重组频率对遗传可操作物种可以被计算，如酵母和果蝇，它们可以进行大量交配，产生上千或上百万后代测定得到统计意义的连锁信息。然而群体较小和有大的代数间隔的物种(包括人类)，这样的杂交是不可能的，数据必须从分析已经存在的家谱中积累。特别是人类连锁作图中存在几个问题。一是人类家庭，特别是含有疾病等位基因的家庭很少能大到产生统计意义的连锁信息。另一个是因为人类交配不是实验所计划的，家谱经常是不完美的，也就是减数分裂未提供重组信息和重组体不能被鉴定。

解决的方法是根据提供的不完全信息计算统计可能性。因为样本小，在人类中遗传距离是以**重组分数**(θ)计算的，它与重组频率一样计算，但以在 0 到 1 之间的数值表示，而不是百分数(也就是一个遗传图谱单位等于 0.01 的重组分数)。从单个家谱中分析减数分裂，计算支持连锁或独立分布的可能性。连锁的统计可能性以“连锁优势”的对数表示，

经常称为**对数值(*Z*)**:

$$Z=\log_{10}\frac{p(\text{两个基因座之间的链锁,假定 }RF=\theta)}{p(\text{两个不连锁的基因座,也就是 }RF=0.5)}$$

一般的策略是计算对于一定 θ 值范围的 Z 和绘制 Z 对应于 θ 的曲线图。对数值>3 被认为是连锁的证据,代表 1000∶1 的优势。最大的对数值 Z_{max} 表示了两个基因座之间最可能的遗传距离。对数值< -2 是不连锁的强有力证据,用负的对数值否定连锁关系称为**排除作图**。

对数值是对数计算的数值,所以它们有可以在家谱中总结数据的优点。尽管这是增加连锁信息的有效方法,但得到的图谱的分辨率比简单生物通过计算重组频率得到的遗传图谱低。另一个增加人类杂交样本大小的方法是测定配子而不是后代的类型,但因为配子不出现人类的疾病表型,所以只有遗传标记而不是真实的疾病可用这种方法作图。

对数值系统的一个缺点是在任何给定家谱的统计连锁可能性是通过假定性状的遗传方式已知的情况下计算的。这适用于研究得比较清楚的孟德尔性状,但对遗传模式不清楚的,可以应用很多稍差一些但不依赖模式的策略,见表 12.5。

表 12.5 人类连锁图谱的其他形式,不需要特定遗传模式。它们描绘了要研究的染色体片段(含有标记),带有特定性状的一些个体都具有此片段

方法	基础和限制
同胞对分析	一对同胞同时患一种特定疾病(因此假定携带同样的疾病等位基因),通过一系列标记测定。他们都具有与疾病等位基因连锁的标记倾向大于 25%(25%的一对同胞会有同时具有任何两个等位基因,即使它们是独立分配的)标记和表型关系越密切,连锁也就越近。
同合子作图	用来对**近亲家庭**(consanguineous families)(近亲结婚的家庭)的隐性性状作图。**同合子**(autozygosity)意味着在特定基因座是纯合子因为等位基因通过**血统鉴定**(来源于同一基因)。在近亲家庭中隐性性状的显现经常表示是同合子,标记的同合子证明连锁。
疾病相关	这种形式的作图是将存在的标记与疾病表型相联系。如果**相对风险(RR)**比 1 足够大或足够小,疾病相关性是明显的。$RR=uz/uy$,w 是患病个体携带标记的频率,y 是未患病个体携带标记的频率,z 是未患病个体不携带标记的频率。

遗传标记 经典的遗传作图涉及确定基因间的连锁,这些基因是形态学特征不同的等位基因。在如酵母和果蝇等被广泛用于遗传作图研究的物种中,存在显示孟德尔遗传的形态学变异的极大群体,携带多个变异的品系可以繁殖并用于基因间直接作图。在人类,遗传作图被用于确定疾病基因的位置:疾病是一种"形态变异",正常和疾病等位基因在每个基因座上是"形态特征等位基因"。然而,在家谱中很少发现,两个疾病基因同时分离,所以通过基因-基因连锁分析直接作图是不可能的。另外,没有足够的孟德尔性状可用于对整个基因组作图。

二倍体物种特异在其他与使用形态连锁标记相关的复杂因素,如未提供重组信息的杂交和等位基因的显性效应-是二倍体物种特异的。减数分裂的产物是配子,需要与另一配子组合产生二倍体个体。这样需要从子代表型推断配子的单元型(亲组合或重组合)。如果来自双亲的配子携带同样的等位基因(未提供重组信息的杂交)或者一个基因

座的等位基因是显性的，就不可能进行作图。如在果蝇中，遗传作图经常用杂合基因型品系和所有研究的基因座上都是纯合的隐性等位基因的**测交原种**(test stock)来进行。这样使每个来自杂合基因型品系的配子的基因型在隐性背景中被确定。然而，人类婚配不能以这种方式控制，所以显性和未提供重组信息的减数分裂存在的问题妨碍连锁分析。

这些问题在很大程度上通过**分子标记**(molecular marker)的发展得到解决——孟德尔性状是广泛分布、易于检测、高度多态性和共显性的，所以大多数个体是杂合基因型，不相关的个体很少携带同样的等位基因，基因型可直接从表型确定(表 12.6)。最有用的分子标记是单一(也是多态的)DNA 序列，可作为探针分离 DNA 克隆和原位杂交确定染色体带，用于物理和细胞遗传图谱中。

表 12.6 遗传作图的分子标记

标记	定义和应用
形态学表型	形态学表型变异。只在具有很多被证明是孟德尔遗传的形态学变异的生物中对作图有用。受显性关系复杂性、非等位基因相互作用和环境效应的影响，人类疾病基因一般是二态表型变异，依赖用其他标记作图。
蛋白多态性	在电泳和等电聚焦时蛋白迁移率的差异。共显性和中等的多态性在数量上有限。很多蛋白的多态性不能通过电泳检测，因为氨基酸替代不改变蛋白的物理－化学性质。同时，基因编码的多态性蛋白本身可能不被作图。
限制性片段长度多态性(RFLPs)	DNA 序列中的产生和缺失限制性位点的差异。是共显性和高丰度的，但经常是二态性，所以不能提供重组信息的减数分裂和杂交十分普遍。很多顺序多态性不能被检测，因为变异位于限制性位点之外。
简单序列长度多态性(SSLPs)串联重复(VNTRs)数目变化。短的串联重复多态性(STRPs)	限制性片段长度或 PCR 产物长度因为小卫星和微卫星随机重复数量的变化形成的差异。因为其共显性和高度多态性是非常有用的标记。小卫星 DNA 主要在端粒区域发现，因此不是理想的基因组范围连锁分析材料。微卫星 DNA 在整个基因组中分布，可以用 PCR 快速测定基因型。三和四核苷酸重复多态性更为常用，因为 PCR 扩增时链的跳格使二核苷酸重复变得复杂化。包含微卫星 DNA 的单一序列 DNA(**多态性序列标记位点**)可以用来鉴定物理图谱上的克隆。
随机扩增的多态性 DNA(RAPDs)	以随机引物扩增的 DNA 片段。丰富，但并非总可以区分在纯合和杂合基因型间，有时重复结果有困难。使用仅限于植物基因组作图。

这样的标记不仅使单个基因可以被作图(**疾病标记作图、基因标记作图**)，通过家谱跟踪进行诊断应用(**基因跟踪**)，而且它们可以用来产生一个覆盖几个物种整个基因组的密集**标记框架**(framework of mark)(**标记－标记作图**)，这种框架可用于进一步的基因作图研究。在小鼠中，密集的标记图谱通过两个品系(如 *Mus musculus* 和 *Mus spretus*)间广泛的回交产生，因为大多数标记类型在两个品系间不同，在回交后代中会产生杂合基因型(因此适合连锁分析)。在人类中，疾病标记作图需要使用疾病分离的家族，而不考虑他们的家谱结构和最终是否适合连锁分析。标记－标记作图不受此方式约束，一系列理想结构的家族在巴黎的 CEPH(Centre pour l'Étude des Polymorphismes Humaines)被集合用于连锁分析。从这些 **CEPH 家族**中建立细胞株，为研究团体提供了稳定的 DNA 来源。

对数量性状的作图　上面讨论的连锁作图策略涉及孟德尔性状或标记，可以通过跟踪家谱或杂交从离散的变异获得连锁信息。然而很多重要的性状是以数量方式遗传的(见

生物遗传和变异),对**数量性状基因座**(QTL)的作图,用这样简单的实验是不能完成的。

如在果蝇这样适合遗传分析的生物中,QTL 是通过费力的繁殖和连锁分析实验被鉴定的。通过很多代的人工选择可以产生出现特定数量性状极端表现的群体,假定每个品系被富集了对某一个极端表型有利的等位基因。使用抑制重组和带有显性筛选标记的平衡染色体,能够从各选择品系中产生携带特定染色体组合的果蝇品系。这些品系可以用来计算它们的数量表型,因此每条染色体所作的贡献可以被测定。用隐性标记的实验可以将 QTLs 定位在特定的染色体带中,在那儿它们处于物理作图的分辨率之中。很多影响刚毛数量的基因就是通过这种方式被作图的,还包括很多涉及神经发育的基因(如 *achaete-scute* 复合物,*Notch*)。

对如猪、牛、水稻和小麦等有重要商业价值的生物,以及人类在自然情况下广泛育种是不可能的。然而,用获得的密集标记图谱,现在可以通过共分离鉴定 QTLs。在农业动物和植物中,QTL 作图的方法涉及两个品系间的杂交或繁殖,这样在特定数量性状上有显著差异,然后计算后代的性状和一系列遗传标记。表型的出现与特定标记的相关性显示了连锁,但这样简单的行为表现 - 亲缘关系相关性不能区分与强 QTL(主要基因)的松弛连锁和与弱 QTL(次要基因)的紧密连锁,因此对定位克隆(参阅)没有帮助。另一种方法,称为**内部作图**(interval mapping),是用类似于优势对数记录分析(参阅)的策略,计算 QTL 在染色体在不同位置存在的可能性。这使 QTL 进一步定位在狭窄区域,进入染色体步移(参阅)的范围之内。

与人类多因子先天性疾病有关 QTLs(**易感性基因座**,susceptibility loci)的作图,也是通过对共分离标记的寻找而进行的。从患者家庭收集家谱数据,患者共有的基因可以通过标记共分离被鉴定。几个涉及胰岛素依赖性糖尿病易感性的 QTLs 通过微卫星标记的复杂的分离分析而得到。在有些情况下,可能分离有特定遗传背景的家族,其数量性状接近于孟德尔遗传。这样的遗传背景可能提供高水平的易感性,在主要易感性基因座特定等位基因的出现或缺失足够引发疾病的开端。在这些情况中,可能可以通过标准的优势对数记录分析或同胞配对分析鉴定 QTLs,前一策略被用于鉴定 *BRCA*1 基因,*BRCA*1 基因是乳腺癌和卵巢癌的主要易感性基因座(见第 18 章癌基因和癌)。

12.9 物理作图

低分辨率物理作图 在哺乳动物和果蝇中,都有根据染色体带型的细胞遗传图谱,物理作图的起始阶段涉及将基因或标记定位在特定的染色体或其中的区域上(表 12.7)。这样的作图策略是低分辨率的,一般将基因座分配到跨越几兆碱基的 DNA 片段中。然而,中期染色体或人工扩增的 DNA 的原位杂交可以使图谱到达 10kbp 的分辨率。

高分辨率物理作图 产生高分辨率基因组物理图谱的策略是将基因组分成许多片段,确定它们的排列,然后分析单个片段分布所包含的基因座,最终测定它们的序列(文框 12.4)。这可以通过制备基因组 DNA 文库(见第 24 章重组 DNA)并利用个别克隆相互覆盖的区域来建立其次序。用这种方法可以产生基因重叠图谱(clone contig map),它能再现染色体的连续区域(表 12.8)。对较小的基因组如细菌基因组,可以使用 λ 取代载体或粘粒载体(见第 24 章重组 DNA)进行基因组作图。虽然在理论上,较大的真核基因

组也可以同样的方式作图，但建立有代表性的文库所要求的个别克隆数目之大超出可能，而且重复 DNA 含量较高导致排序出现问题。因而作图的起始阶段要求容量较大的染色体，如**酵母人工染色体**(yeast artificial chromosome，**YACs**)（部分常用的人工染色体的性质在表 12.9 中进行比较）。一旦产生了这样的图谱，个体 YACs 可以被亚克隆到粘粒中进行更精细的定序。最后，粘粒的插入片段可以用限制性内切酶分成更小的片段，亚克隆后进一步鉴定。

表 12.7 低分辨率物理作图的技术

作图策略	基 础
定位在单个染色体上	
体细胞杂交	**体细胞杂交**通过融合不同种类的培养细胞来完成，例如用聚乙二醇处理。在人类基因的作图中使用了啮齿动物/人类杂交细胞。一般情况下，起始的杂交是不稳定的，大部分人类染色体不能复制，最终产生具有一整套啮齿动物染色体与一条或几条人类染色体的稳定杂交细胞。获得这类杂交细胞的库或称**杂交细胞组**（hybrid cell panel)就可以通过 PCR 或杂交实验将特定的人类 DNA 片段清楚地定位于人类染色体，在某些特例中可以通过测定基因产物而定位。将人类微细胞与啮齿类细胞融合可产生**单染色体杂交**细胞（只包含单个人类染色体的杂交细胞)，这样使用包含 24 种细胞株的组就可以清楚地定位人类 DNA。**微细胞**（microcells）是类似细胞的颗粒，它的较小的核中包含单个染色体，核外围绕着少量的胞浆和细胞膜，它们由离心后的有丝分裂延长终止产生。
剂量作图	对有多重拷贝特定染色体的细胞株或体细胞杂交细胞进行分析，通过定量 PCR，杂交或产物的表达监测到的剂量，可以定位超表达染色体上的基因。
定位到染色体亚区域上	
缺失或易位作图	对包含易位或缺失的接受染色体的杂交细胞组进行分析。这种物理作图的方法包括利用杂交或 PCR 对 DNA 序列进行测定，或对基因产物进行分析。同名的细胞遗传学作图技术包括将某种表型与细胞遗传可见的染色体重排相对应而获得基因的位置。
原位杂交	将一核酸探针与染色体杂交获得在特异染色体带的定位。传统的体外杂交使用放射性探针，现已被使用非放射性的荧光探针的**荧光体外杂交（FISH)**取代。除了它的速度和效率，这个方法的另一个优点是可以同时使用具有不同的荧光生色基团的探针根据颜色的不同来鉴别不同的靶位点，从而可以确定基因次序（又参见染色体着色）。对中期染色体的 FISH 给出 1－10Mbp 的解析度；然而同样的技术也可应用于间期染色质，或应用于人工延长的染色质纤维（**直观体外杂交，DIRVISH**)和裸 DNA（**DNA 纤维杂交**)，解析度小于 10 Kbp。去折叠 DNA 的广泛成环要求包括延伸纤维 FISH 的作图工作需通过对结果的统计分析来印证。

物理作图速度与效率通过网格库(gridded libraries)的引入有了较大的提高。也即文库克隆（例如全基因组的 YACs 或从特异的 YACs 而来的粘粒)被挑出并置于多孔板的单个孔中培养。然后，不像常规的筛库方法那样将克隆转到滤膜上（参见噬斑移动，克隆印迹)，而是以网格状排列将克隆点到滤膜上，这样每一克隆可以精确地对应。这一技术减少了在克隆辨别中的混淆还可以使网格库进一步分散到其他库中进行相关工作。

表 12.8　用来产生基因组克隆重叠的技术

技术	原理
染色体步移	在理论上,可以使用随机克隆筛选基因组文库获得相互覆盖的克隆,然后将这些克隆作为探针以同样的方式将获得更多的相互覆盖的片段。实际上克隆与克隆之间的步移由于重复 DNA 的出现而复杂化,特别当使用如 YACs 这样的大容量载体时。重复 DNA 可以通过与已知的重复序列预杂交而抑制(例如人类基因组中的 *Alu* 顺序)。而且实际上并不使用整个克隆而使用标记的末端片段为探针。与 YACs 有关的另一个问题是嵌合性(表 12.9),因而 YAC 的步移往往要求染色体制备的 FISH 实验为补充。
杂交作图	用一套随机克隆筛选文库来辨明相互覆盖的克隆。从余下的阴性克隆中,选择另一套随机克隆并重复筛选过程。重复这种策略直到所有的克隆都被杂交。
寡核苷酸作图	寡核苷酸与一系列粘粒克隆杂交。与同样的寡核苷酸杂交的粘粒应该是相互覆盖的。也可以将寡核苷酸设计为处于共有基序周围来获得序列信息。
限制性片段指纹	用一组限制性内切酶消化克隆,这样对每个克隆都产生一限制图谱。个别克隆的图谱相互比较获得相互覆盖的区域。
顺序标记位点(STS)	通过测试随机亚克隆的 DNA 片段如测试它们在基因组 Southern 印迹中获得单个条带,或用 PCR 获得单个产物的能力来鉴定单一序列,它们被称为序列标记位点 (STS)。用这些标记物来测试整个基因组,包含同一 STS 的两个或多个克隆必然相互覆盖。通过测试 cDNA 而非基因组克隆可能可以加快潜在 STSs 的鉴别,因为在 cDNA 中更可能包含单一的 DNA 序列。
重复 DNA 指纹	包含很多重复 DNA 的哺乳类基因组可以用来获得克隆重叠。重复 DNA 指纹包括 YAC 或粘粒用限制性内切酶消化,并用重复 DNA 探针作 Southern 印迹,例如 *Alu* 元件特异性探针。具有同样的带型的克隆可能相互覆盖。类似的基于 PCR 的技术使用单个 *Alu* 特异的引物扩增头对头 *Alu* 元件之间的基因组 DNA,通过电泳对扩增产物进行分析,具有同样的带型的克隆可能相互覆盖。

表 12.9　基因组作图中使用的高容量人工染色体克隆载体的比较。人类人工游离染色体(human artificial episomal chromosome, HAEC)是一类发展中的 MACs,它来自 Epstein-Barr 病毒

载体	系统	克隆容量	评价
YAC(酵母人工染色体)	酵母着丝粒,复制起点和端粒	200kbp～2Mbp	嵌合发生率高(在基因组中不相连的连续片段) 不稳定(自发缺失) 很难与酵母染色体分离
PAC(P1 人工染色体)	细菌噬菌体 P1	100～300kpb	稳定,但在细菌外难以维持除非整合到寄主 DNA 中
BAC(细菌人工染色体)	F 质粒	300kbp	
MACs(哺乳类人工染色体)	基于 Epstein-Barr 病毒的游离载体	>300kbp	

限制图　限制性内切酶可用来产生基于限制位点位置的物理图谱(**限制图**, restriction maps)。图的解析度依赖于 DNA 片段中限制位点出现频率,而这又由限制位点的大小和碱基组成所决定。具有 4～6bp 识别位点的酶可用来产生小 DNA 分子如质粒,PCR 片段

和 λ 插入片段的限制图。具有较大的限制内切位点(8～10bp)和/或识别未被充分代表序列如哺乳动物基因组中的 CpG 的内切酶称为稀有切割者。可以用这种酶来获得整条染色体的限制图谱,虽然产生的 DNA 片段必须用*脉冲场凝胶电泳*(参阅)等方法分离。分离获得的限制性片段可以通过杂交（参见 southern *印迹*)或 PCR 来测试所携带的标记。使用一组限制性内切酶使片段定序并相互重叠。对于限制性内切酶在分子生物学中的使用的讨论,见重组 DNA(本章也提供了一个质粒载体的限制图的例子)。

真核基因组的作图与鉴定　　在较高级真核生物的大基因组中,大部分 DNA 不表达,它们中很多是重复的。因此有必要设计基因鉴定的特异性策略。

在理论上,基因可能由于它的序列与以前鉴定的基因保守,或它在基因组中具有独特的结构,或因为它表达产生 RNA 转录本等而被鉴定。上述策略都被应用(表 12.10)。选择脊椎动物基因组中富含基因的克隆包括 CpG 岛作探针来确定基因的 5′端。还可以使用河豚鱼的基因组克隆,其基因组的复杂性大于 90%。这些策略都有它们的缺点:估计的 70 000 个哺乳类基因只有一半与 CpG 岛有关联,也并非所有的基因具有与河豚基因足够的同源性可以以杂交鉴别。

表 12.10　基因组 DNA 中基因鉴别的方法。鉴别特异基因的策略参见*定位克隆*。

方法	策略
利用序列保守性	
种间同源	种间基因往往具有保守性,而非编码 DNA 则不然。因而可以用亚克隆对不同种属的基因组作 southern 印迹,来显示保守的基因序列(参见 zoo *印迹*)。
数据库同源搜索	来自一亚克隆的序列可以与序列数据库中的序列相比较。可能会发现与已克隆基因的同源区域。
河豚鱼比较筛选	对低复杂性的脊椎动物基因组,与河豚鱼基因组克隆的杂交可帮助鉴别基因,因为河豚鱼基因组的复杂性很高。
外显子预测	基于开放读码框与剪切顺序的出现用计算机程序来预测假想的外显子位置。
利用独特的结构	
CpG 岛的鉴别	在高级真核生物中包含 CpG 岛(参阅)的区域往往标志着基因的 5′末端,可以通过杂交或识别位点具有 CpG 基序的限制性内切酶图谱或 PCR 来辨别。
利用基因的表达	
表达杂交	标记基因组探针的 northern 印迹杂交来辨明转录区域,在 cDNA 库中分离表达基因。
外显子捕获	基因组克隆插入到表达载体中两个外显子之间的内含子中,构建完成后转化到细胞内。如果基因组克隆中包含外显子,剪切会产生具有三个外显子的成熟转录体(两个载体外显子和中间被捕获的外显子)。如果不然,转录本只包含两个外显子。细胞中分离的 RNA 用 RT-PCR 的方法获得此外显子(**外显子扩增**)。
cDNA 选择与俘获	cDNA 与基因组克隆(或固定于一滤膜上或在溶液中但用生物素标记以利纯化)杂交。基因组 DNA 与 cDNA 的异源双链通过洗涤或链霉亲合素纯化,cDNA 用 PCR 扩增和鉴定。扩增 cDNA 可经历几轮基因组杂交以富集阳性序列。

基因鉴定的另一种方法是特异性的克隆并鉴别表达 DNA 而不是从基因组 DNA 中筛选。这可以通过对 cDNA 的广泛鉴别或外显子捕获或 cDNA 俘获等策略来进行（表

12.10),但这些方法的缺点是不能提供基因结构与调节信息,而且瞬时表达基因或只在特异细胞/以最低水平表达的基因会被漏掉。集中研究哺乳类基因组的表达序列,有可能将基因组中<5%,但承担大部分功能因而最具商业与医药价值的序列定位。较长长度的cDNA克隆并继以短的cDNA片段的随机测序产生200~300bp的**已表达序列标鉴**(expressed sequence tags, **ESTs**)库,这种标志可以通过杂交和PCR定位到物理染色体图谱上。与其他基因鉴定的方法协同进行,这种工厂化的基因鉴定方法可能使所有的基因的染色体位置被确定(**转录图谱**),从而提供完整的基因图。ESTs项目已开始鉴定人类不同组织中的cDNA,其中大部分信息是冗余的。然而产生的序列的绝对数目(>200000)可以足以定位大部分基因。EST可以用放射杂交定位到染色体上,或是重叠图的个别YACs上以产生全长cDNA重叠顺序(**EST步移**)。

虽然EST鉴定可以鉴别很多表达序列,它几乎不能提供定量信息。最近已发展了几种技术以对多个基因座的基因表达同时进行定量分析(例如SAGE,寡核苷酸芯片,参阅)。这些方法,以及通过突变来确定功能的相关方法和利用双杂交系统研究基因产物的相互作用的方法,一起组成了快速延伸的**功能基因组学**(functional genomics)的领域。它将开发基因组测序获得的信息并利用它在整个基因组范围内将功能与DNA序列联系起来。进一步的讨论见第22章蛋白质:结构,功能与进化。

比较基因组作图　　比较基因组学是基因组科学的分支,它涉及不同种类基因组间的比较。这种比较有两个目的:提供基因与基因组进化有关的信息(即在基因组水平了解种间的相同与不同之处),并通过**比较**或**同线性作图**(comparative or synteny mapping)为基因克隆提供帮助。脊椎动物种间的比较作图特别有价值,因为它可能为人类的遗传疾病提供新的动物模型和新的治疗方法,同时也可提供脊椎动物进化模式有关的信息。河豚鱼基因组是特别有用的比较作图工具,因为它的堆积程度与基因密度高,在鱼与其他的脊椎动物中谱系往往是保守的,但河豚鱼基因组中的基因更容易分离,因而可用作探针来监测保守的哺乳类基因。

脊椎动物的比较作图通过不同水平的保守染色体片段来描绘。低解析度的比较作图可以用**染色体着色**(**zoo-FISH**),即从一个种类的单个染色体上分离得到的DNA被扩增,用荧光探针标记后,与另一种类的中期染色体进行原位杂交。人类与猫的染色体之间的比较显示了广泛的同线性,在某些例子中观察到整条染色体的保守,而在人类与小鼠中保守区域要小得多[X染色体特别强烈地保守,因为对X偶联基因座有剂量补偿效应的限制(参阅)]。在更精细的尺度中,比较作图可以显示在同线性片段的标记之间的保守联系。用于比较研究的标记类别为基因(在种间差异很小的基因),而不是在谱系研究中使用的高变的微卫星标记(种内具有多态性,在种间保守程度低)。通过比较测序获得的功能性DNA序列的保守性可帮助鉴别基因和基因外的调控元件。哺乳动物中的比较作图由**锚参比基因座**(anchor reference loci)的使用而增强。它们位于染色体片段中,表现了所有构建中的哺乳类基因组图谱的保守联系(这种片段如**SCEUSs:最小的保守进化单位片段**,smallest conserved evolution unit segments)。在SCEUs或其他区域鉴别的单一序列(序列标记位点)可以用来产生标记图谱,使所有将获得的哺乳类基因组图谱相互联系。在所有的哺乳动物中保守的序列标记位点被称为CATS(**保守锚标记序列**,conserved anchor-tagged sequences)。

文框 12.1 决定基因组性质的复性动力学

Cot 分析 在通过测序进行基因组分析成为可能之前,**复性动力学**(reassociation kinetics)(分析单链核酸在溶液中退火行为)被用来研究基因组性质。尽管这种技术现在主要是具有历史意义,但其原理对理解基因组结构和核酸杂交(参阅)还是有用的。双链 DNA 可以通过加热被变性或熔解(分成单链),如果逐渐降温,会重新组合(复性、退火)形成双链分子。因为反应有两条链,所以单链 DNA 在溶液中重新组合遵循二级动力学,发生的速度可以以公式 12.1 表示,C 是在时间 t 时单链 DNA 的浓度;k 是**复性速率常数**。一定比例的单链 DNA 分子在任何时间都存在,起始浓度 C_0 可以通过积分确定,见公式 12.2。这证实了 C_0 的产物和 t 是控制复性速度的参数。

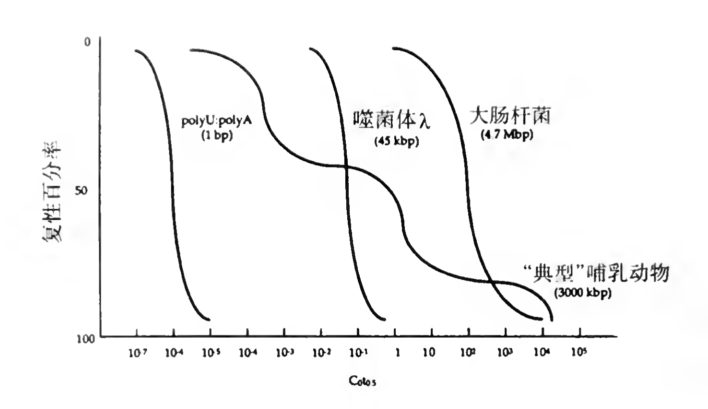
$$\frac{dC}{dt} = -kC^2 \tag{12.1}$$

$$\frac{C}{C_0} = \frac{1}{1+kC_0t} \tag{12.2}$$

一半 DNA 复性的点($t_{0.5}$)被选作参考点。在这点上,$C/C_0=0.5$,通过对公式 12.2 的重排,表现出 $C_0t_{0.5}=1/k$。$C_0t_{0.5}$称为 **Cot 值**,与基因组复杂性成正比。这是因为复杂性增加,任何单个顺序相对浓度下降,需要更长的时间找到互补的链。复性反应因此需要更长的时间达到中点。

Cot 曲线 从基因组 Cot 分析得到的数据经常以 $\log_{10} C_0t_{0.5}$对复性 DNA($1-C/C_0$)作图,产生 **Cot 图**或 **Cot 曲线**。对细菌和病毒的简单基因组,复性发生在 Cot 值的两个数量级范围,Cot 曲线在其长度的 80%是线性的。随着基因组复杂性的增加,$C_0t_{0.5}$增加,曲线右移。大肠杆菌和 λ 噬菌体的 Cot 曲线与人工的复杂性为 1 的多聚尿嘧啶/多聚腺嘌呤在下面一起表示,进行类似分析的真核生物基因组显示出在更大的 Cot 值范围内进行复性。真核生物 Cot 曲线经常分成三个重叠的曲线,代表了不同顺序复杂性的基因组片段,有时称为**快速、中间和缓慢成分**,对应于高度重复、中等重复和单一顺序 DNA。缓慢成分是对真实基因组复杂性的最好预测,因为大多数基因在单一顺序 DNA 中发现。一部分 DNA 几乎立即退火。这些**零时间结合的** DNA 也被称为**反折 DNA**(snap-back DNA)或**折返 DNA**(fold-back DNA),因为其代表了可以通过分子内碱基配对杂交的二元对称区域。标准哺乳动物的 Cot 曲线比上述简单基因组更复杂。

Rot 分析 复性动力学还被用来研究 RNA 成分的性质和丰度。RNA 在溶液中与互补 DNA 复性,遵循类似于 DNA 复性的动力学。反应的驱动参数是起始 RNA 的浓度和时间,称为 R_0t 或 **Rot 值**。RNA 与 cDNA 的重新组合用来确定 RNA 的复杂性,也就是细胞中不同 RNA 分子的代表性。这类重新组合实验产生跨越几个数量级的宽 Rot 曲线,一般分解成三个成分:高丰度成分,中等和稀少成分。**高丰度成分**在低 Rot 值杂交,因为其包含少于 50 个 mRNA,经常称为**简单成分**。这些成分在每个细胞中出现高达 10000 次,达到整个 mRNA 群体的 50%。高丰度成分经常代表了组织特异性转录本:例子包括 α-和 β-珠蛋白、肌动蛋白、肌球蛋白和白蛋白的 mRNA。**中等丰度成分**在 10^{-2}到 10^2 Rot 值杂交,包含 100~1000 个转录本,每个出现几千拷贝数。**稀少或稀有成分**在高 Rot 值杂交,经常称为**复杂成分**,因为其包含了上万个转录本,每个出现少于 100 次。大多数管家基因在这种成分中发现。



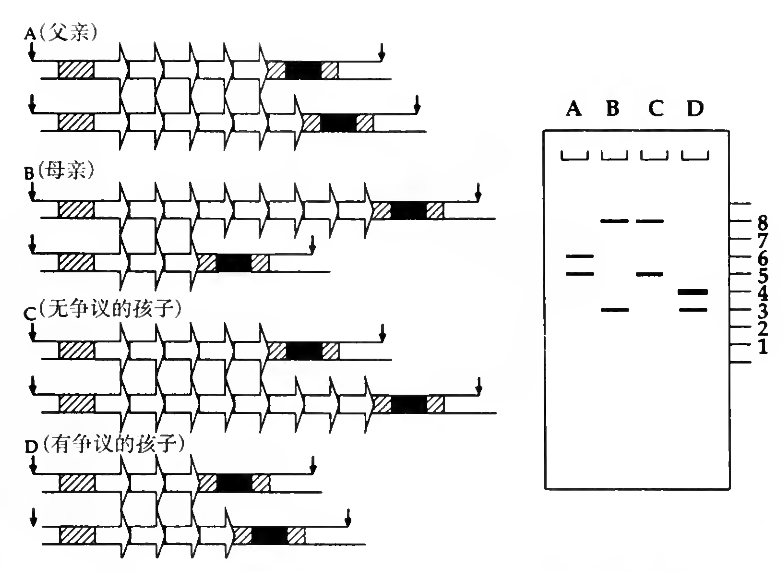
驱动物和追踪物　　特殊复性成分的行为可以通过在复性实验中包含少量放射性标记的材料来研究,这样的材料称为追踪物。反应的动力学由过量出现的复性成分控制,这样的成分称为**驱动物**(DNA 出现大于 RNA 的反应称为 **DNA 驱动反应**,遵循 Cot 曲线,相反的 **RNA 驱动反应遵循** Rot 曲线)。一个在 DNA 复性反应中的 RNA 追踪物可以鉴定表达的 DNA 成分,这类实验用来确定单一顺序 DNA 中的大部分基因。在需要鉴定复性反应中不杂交的成分时,反应可以进行达到饱和,可以观察到 Cot 或 Rot 曲线的平台。这样的**饱和动力学**实验可以被用来如确定不出现在特定 RNA 群体中的 DNA 比例。

文框 12.2　DNA 分型

DNA 分型的基础　　**DNA 分型**(DNA typing)或 **DNA 分布图**(DNA profiling)涉及用小卫星 DNA (VNTR)产生 DNA 片段组,以电泳分离时,提供任何个体的独特模式(这样的分型有时称为 **DNA 指纹**)。小卫星 DNA 是高度多态性的(每个位置重复单位的数目),而在基因组中有很多小卫星 DNA,倾向于分布在亚端粒区域。如果足够的位点被同时分型,不相关的个体极不可能产生相同的分布图,但因为小卫星是以孟德尔性状传递的,相关个体会有类似分布图,并且相匹配的 DNA 片段数目与对应于它们亲缘关系的紧密程度呈正相关。

应用　　DNA 分型产生个体特异性 DNA 分布图的能力可应用在犯罪研究中。DNA 可以从犯罪现场的组织和体液中提取(经常是血液、精液或毛发),然后与怀疑对象取得的对照样品比较。同样的,DNA 也可以从动物和植物中获得,与保存的参照比较确定它们的起源,如在被偷盗的濒临灭绝鸟类和它们的蛋的情况中。相关个体 VNTR 等位基因分享的倾向可以加以利用。这可帮助确立亲子关系(见下面的例子),证实家谱或显示个体的相关性(如被迁移打断的)。

DNA 分型方法学　　原先的 DNA 分型方法涉及用限制性酶剪切 DNA,通过基因座特异性探针进行 Southern 杂交(参阅)分型。限制性片段大小将依赖于每个小卫星重复单位的数量。这些技术需要相对大量的新鲜(未降解的)DNA,而法院证供的证据经常只有少量,且经常是陈旧而降解的。这些问题在一定程度上通过扩增微卫星重复聚合酶链式反应(参阅)得到解决。**PCR 分型**方法类似于分布图,但可以应用于微量样品(如干了的一滴血,一根毛发),并可容忍一定程度的 DNA 降解。然而,必须小心避免外源的污染,按常规样品和对照常规在不同的实验室进行。



图中显示了一个简单的 DNA 分型，用于父子关系的争议。父亲的 VNTR 基因座(A)、母亲(B)和两个小孩(C,D)在图中显示。小孩 C 的父子关系无可争议，但 A 怀疑他不是小孩 D 的父亲。从血液中的得到的 DNA 用限制性酶剪切(小箭头表示跨越 VNTR 顺序的限制性位点)，通过电泳分离，以对应于不变区域(深色框)的探针 Southern 杂交。每个等位基因产生不同大小的电泳条带，依赖于重复的数量(以数字表示)。分布图显示小孩 C 从每个亲代遗传了一个 VNTR 顺序，小孩 D 从母亲遗传了一个 VNTR 顺序，另一个等位基因不是从 A 中得到的(以黑色带表示)。A 假定他不是小孩 D 的父亲，因此可能是正确的。

文框 12.3 连锁作图精确性的限制

限制的总结 当考虑到非常大或非常小的基因座之间的距离，连锁作图本身是不准确的，这是因为：

(1) 不考虑基因座之间的距离，最大的重组频率是 50%，这是因为交换的统计分布考虑到四条染色体是等价的；

(2) 多次交换的影响当基因座之间距离变大时导致对重组频率的低估变大；

(3) 当基因座之间的距离变得小时、非交互重组对重组频率分析的影响。

用重组频率的连锁作图在短距离是精确的(也就是多次交换不会发生的地方)，但在被研究的两个基因座距离太近，重组会将它们包括在一个异源双链 DNA 的片段中。

(4) 不同重组频率的区域和重组热点和冷点的存在，意味着遗传和物理距离间关系不是不变的。重组频率在性别和物种间也有变化。

最大重组频率 不同染色体上基因座独立分配，重组频率是 50%，因为亲本染色体的重组以重组组合频率进行，这是因为中期板同源配对的随机方向。对于同一染色体上的基因座，独立分配是不可能的，但重组单元型可以通过交换产生。随基因座之间距离的增加，交换的机会也增加，但重组频率不会超过 50%，因为即使距离是远到交换一定会发生，一个交换涉及二价染色体的两个染色单体和只有减数分裂的一半产物是重组的。如果考虑双交换，不同类型的双交换的分布也确保最大重组频率是 50%：只有涉及所有四条链的双交换产生四个重组合(100%重组)的染色体，但

只涉及两条链的双交换统计上同样发生的，这样产生四个亲组合染色单体(0%重组)。涉及三条链的双交换产生两个亲组合和两个重组合染色单体(50%重组)。这样整个重组的频率是50%。

多次交换　遗传距离对应于物理距离的线性范围是很短的(大约15个图距单位)。随基因座之间距离变大，通过重组频率确定的距离偏低于标记间真实的遗传距离。这是因为**多次交换**的影响。

在杂合基因型标记间单个交换将亲本基因型变成重组基因型，通过交配产物的分析计算这种改变的基础上，连锁作图才起作用。到某一点之前，基因座之间增加的距离增加了基因座单交换的可能性，使重组频率准确预测物理距离。然而，随着进一步分离，开始发生双交换。这样的情况可以被计为两次单交换，但在两个基因座交配中，双交换类型是不能与亲组合区分，就不能在计算中被扣除。因此交换事件的计数比真实值低，重组频率和基因座之间的距离也被低估了。

基因座距离的进一步增加，更多次交换可能发生。然而，三交换会被计为重组合，四交换类型是亲组合的，这因为基因型无法区分。遗传作图的基本弱点是任何偶数交换被分型为亲组合，而任何奇数交换是重组合。最后，基因座之间的距离变得如此之大，以致产生偶数交换的可能等同于奇数，重组频率为50%。在这点上，即使在同一染色体上的基因座的行为如同它们是独立分配的。

多点交换和作图功能　在原则上，遗传距离的低估可以通过测定多次交换类型或用被低估的一个预测真实距离来校正。

在遗传易操作的物种中，对多次交换的测定可以通过包括更多基因座的杂交进行，如在经典的果蝇三点测交中，这样作图的区域分成较小的基因座距离。在人类中，疾病基因对标记框架的多点作图比两点优势对数分析更优越，因为有了多种不同的标记，未提供重组信息的减数分裂的机会就减小了。用两个以上基因座也能确定基因在染色体上排列顺序。在有些真菌中(如曲霉菌，**囊霉菌**)减数分裂的四个产物保留在一起，在称为**子囊**(ascus)囊状结构中作为**四分体**(tetrad)存在，这样，可能得到校正双交换的连锁数据，因为每个减数分裂染色体的产物可以被鉴定(**四分体分析**)。

对遗传图距的校正通过作图函数来实现，作图函数是重组频率和遗传距离之间的数学关系。有三种类型的作图函数，Haldane 的、Kosambi 的和 Ott 的。Haldane 的函数是最简单，因为它假定交换是随机分布的，没有相互干扰(见下)。它可以下面的公式表达，d = 遗传距离，r = 重组频率：

$$d=\frac{-\ln(1-2r)}{2}$$

干扰　当一个重组发生时，它是否会影响另一个的起始？**干扰**(interference)表述的就是这样的效应，**正干扰**是一个交换抑制另一个，而负干扰是刺激另一个。在果蝇中，通过观察特定三点杂交中双交换的数量是否是每个单交换类型频率所预期的，可以来预测干扰。乘积定律规定两个事件同时发生的可能性等于每个单个事件发生可能性的乘积。双交换预期频率因此是观察到的单交换频率的乘积干扰按下式计算的，I 是干扰指数，c 是**并发系数**：

$$I=1-c$$

而

$$C=\frac{\text{观察到的双交换频率}}{\text{预期双交换频率}}$$

在真核生物中，一般观察到重组是正干扰的，也就是一个交换抑制附近另一次交换的起始。

然而,非常接近标记间的重组显示了负干扰的证据,也就是几个交换呈现出成簇现象。事实上这是杂合双链 DNA 加工产生的假象,可以用*基因转换*(参阅)而不是链互换来解释。

区域重组频率变化、热点和冷点　　连锁作图依赖于交换的随机分布,但在基因组中存在区域变化以及重组的促进位点(**重组子、重组遗传元件、重组热点**)和抑制位点(**重组冷点**)。

在人类中,在每个染色体中都有重组频率的区域变化。交换在染色体端粒区域比着丝粒发生更频繁。雌性比雄性重组频率更高一些,而雄性在 X:Y 对的拟常染色体区域显示出专性的交换,所以重组频率经常是 50%。值得注意的是 Y 染色体没有减数分裂图谱,因为它不涉及交换事件,尽管它可以通过放射性杂交作图得到遗传图谱。

重组热点经常是内切酶靶位点,因为剪切提供了起始*同源重组*(参阅)的单链 DNA。这样的位点包括大肠杆菌中的 *chi* 位点和其他细菌中的相关位点。重组热点引起对遗传距离的高估,因为如果重组倾向于在特定位点起始,这些位点两侧基因座的重组比平均水平频繁。然而,纵观整个基因组,重组热点相对平均分布。仅在考虑短距离时它们影响明显,导致酵母杂交中的**极性现象**,在这种现象中,涉及基因转换的不同等位基因有不同的频率,依赖于从交换起始位置*分支迁移*(参阅)的范围。

重组冷点经常是同源 DNA 不能有效联会的位置。这经常发生在有不育阻碍的位置,如一条染色体倒位或易位杂合型,在断裂点处联会被阻止。这种位点两侧的基因座显得比它们实际上距离近,这是因为它们之间的重组频率低。复杂的重排异构体缺乏联会能力可被利用来阻止重组。

文框 12.4　DNA 测序

DNA 测序方法　　1977 前,确定 DNA 中碱基顺序(**DNA 测序**)是费力的工作,且只能应用于如 tRNA 的小分子。在这时发展了两个快速大量 DNA 测序的技术,两个都涉及产生一套重叠的长度相差一个碱基的 DNA 片段。四个测序反应平行进行,每个产生确定末端碱基的重叠的片段。这些反应的并排电泳使顺序通过电泳胶或放射自显影读出。

Maxam 和 Gibert 测序　　包括用修饰特定碱基的化学试剂降解限制性片段。**Sanger 测序**方法包括 DNA 合成,每个反应包括少量四种**双脱氧核苷酸三磷酸(ddNTPs)**的一种。它们是**生长终止剂**(telogens),也就是引起链终止的物质,因为它们缺乏延伸所需的 3′羟基基团,因此这种方法经常称为**双脱氧方法**或**链终止方法**。Maxam 和 Gilbert 测序起初是最常用的,因为它能用限制性酶和通常的实验室试剂进行,而 Sanger 方法需要特定的试剂和 M13 载体。随着噬菌粒(phagemid)的发展和优良试剂如双脱氧核苷酸的商业化,Sanger 方法变得更普遍。它最适合自动化大量测序计划,并且大多数现在进行的一般测序也是用这种方法。Maxam 和 Gilbert 测序对有些特殊的应用如 DNase *足迹法*(footprinting)(参阅)还是有用的。

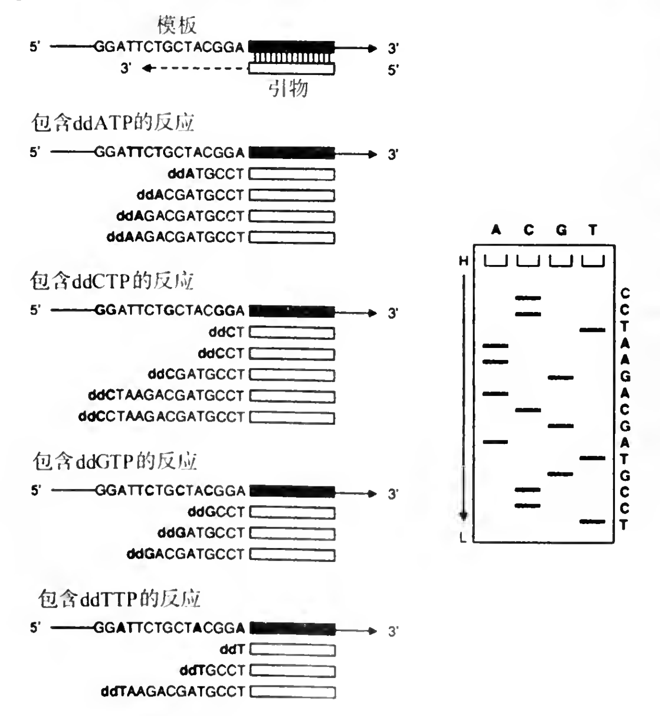
最近发展了几种新的测序方法,其中两种可能在将来会被应用。第一个是**扫描隧道显微镜**和类似原理的**原子力显微镜**可以对 DNA 分子表面结果作图,随技术改进有可能分辨单个碱基。第二是**杂交测序**,是用系列的固定化寡核苷酸产生杂交图谱。杂交测序要求在寡核苷酸自动化合成以格子分布方面的提高,如所有可能的十聚体需要合成和定位超过 50 000 分离的分子。在将来,可在小的芯片上分布上千个寡核苷酸用于杂交,通过计算机直接分析图谱,快速确定顺序。

Maxam 和 Gilbert 测序　　在这种技术中,进行四个反应,末端标记的限制性片段与修饰或去除特定类型碱基的试剂作用。硫酸二甲酯甲基化鸟嘌呤,酸去除任何嘌呤肼修饰所有的嘧啶,肼和氯

化钠特异修饰胞嘧啶。修饰的碱基被哌啶去除,链在脱碱基位点处剪切。试剂在只引起每条链修饰一个位置的浓度下使用。这样产生有同样标记末端的和不同碱基特异性非标记末端的重叠片段系列。四个反应的产物在电泳胶上并排电泳,顺序通过放射自显影读出。

标准链终止测序 在这个技术中,引物与单链 DNA 退火,通过 DNA 聚合酶延伸——T7 DNA 聚合酶的高活性重组形式称为**测序酶**(sequenase)经常被使用。同位素标记的核苷酸在引物延伸中掺入,有四个反应,每个都包含少量四种双脱氧核苷酸中的一种。这样每个反应产生一套重叠标记产物,从测序引物开始,在特定碱基终止。这些通过邻近泳道的电泳,使顺序由放射自显影读出(见图)。起先,这种技术需要以 M13 载体为基础的单链模板。这样的模板会得到最佳的结果,但也可以进行**双链测序**,通过对质粒等的 dsDNA 模板先变性。测序引物只与一条链退火,所以只有一条链被用作模板。

近来对链终止测序的改进 最近对链终止反应的很多改进是从自动化过程和增加得到顺序信息而获得。**多路测序**是很多克隆在同一反应中测序,并在同样的泳道上电泳:每个克隆并非通过插入标记测序,而用独特的引物进行,引物用于 Southern 杂交确定对克隆特异的测序条带。**染料终止测序**使用荧光染料标记的双脱氧核苷酸。用四种产生不同波长的光染料,每个对应于一种碱基。这使四个反应在一个泳道上进行,顺序由电泳过程中凝胶末端的探测器读出(**实时测序**)。这不仅增加了流通量,且因为信息直接从探测器到计算机,减少了记录顺序的误差。**循环测序**是链终止测序和 PCR 的杂合。用双链模板和单链测序引物,但反应用热稳定 DNA 聚合酶(参见第 21 章聚合酶链式反应)通过热循环进行。循环测序降低了由模板二级结构引起的假象,因为产物以线性方式积累(参见不对称 PCR),只需要少量模板。循环测序比标准测序精确度低一些,因为热稳定聚合酶如 Taq DNA 聚合酶是差错倾向较高。



基因组计划中的测序策略　对大量测序计划，如基因组测序，大的基因组克隆经常随机亚克隆到噬菌粒载体，产生很多相互重叠的克隆。这些克隆被随机挑选和测序（**鸟枪测序**，shotgun sequencing），顺序信息输入计算机，由计算机检测重叠和将插入片段排列成连续的分辨率为一个碱基的图谱。任何存在的孔隙通过在原始基因组克隆中的**引物步移**（priner walking）补平，引物步移是引物根据接近孔隙的已经测序顺序设计引物，得到缺失的顺序。

顺序信息的操作和储存　生物信息学是指快速发展用于储存、分析、比较和组织 DNA 和蛋白顺序信息的计算机技术。有三个并列的数据库：EMBL（由欧洲信息研究所发布），Genebank（由美国国家生物技术中心发布）和 DDBJ（日本的 DNA 数据库），包含了主要的 DNA 顺序信息，并随新序列被发表不断更新。特殊的数据库也有提供，用来保存快速扩展的基因组计划中的表达顺序标记（见正文），及显示基因组作图，基因顺序和表达分布的模式生物数据库。数据库通过互联网对世界范围的研究科学家开放，并可以和多种精妙的作图、顺序排列和预测蛋白顺序、同源性和功能的分析程序共同使用。现在，大多数顺序信息是免费开放的，但在基因组科学中不断增长的商业价值，不可避免造成垄断的危险性，如医药公司和农业公司会控制基因组顺序来保护它们的利益和投资。

（陈　淳 译）

第13章
可移动的遗传因子

基本概念和定义

- **可移动的遗传因子**(mobile genetic elements)是指可在同一细胞中基因组内或一个细胞的基因组间从一个位点移动到另一位点的 DNA 片段。它在原核生物和真核生物中均有所发现,由于结构、移动机制、分布、移动自由度、自主水平的不同而分为许多种类。
- 虽然一些因子的移动具有明确的细胞功能,但大多数因子被认为是自私 DNA 或偶然发生移动的 DNA。在高等真核生物中,很大一部分的基因组是由有功能和“假”(ghost)转座因子组成(由突变而失活)。一些细菌的移动因子发展了一种细菌在宿主细胞的共生关系。可移动的遗传因子在基因组进化中起着重要作用。
- 可移动的遗传因子根据它们独立的水平可大致分为三类。第一类为**附加体**(episome),如噬菌体 λ、F 质粒。此类因子能以两种形式存在－或整合在宿主基因组内,或作为染色体外的自主复制子。第二类为**暗盒**(cassette)[1](例如酵母的交配型盒,锥虫 VSG 盒等)。此类因子的移动由细胞控制,并介导特异性的基因组重排。附加体和盒式元件的移动均利用同源重组或位点特异性重组(参阅),即对靶位点的选择很严格,一般只是一个位点。第三类称**转座因子**(transposable elemement)(如 P 因子、反转录病毒等)。转座因子在宿主基因组外不能复制,但可在宿主基因组内调控自发移动。转座因子通过**转座**作用发生移动,此作用无需转座因子与靶位点同源性而形成重组,所以在靶位点的选择上有一定自由度。转座因子是可移动的遗传因子中最大的一类,它在结构、转座机制、自主水平方面呈现多样性。
- 有些可移动的遗传因子是病毒。噬菌体 λ 是病毒附加体的例子。噬菌体 Mu 是转座因子的例子——它必需整合到宿主基因组中才能复制。真核生物反转录病毒也是转座因子的例子。
- 根据转座机制不同,转座因子可分为两大类。**I 类转座因子(反转录因子,** retroelement)它在转座过程中利用 RNA 来介导——整合因子先转录,再反转录为 cDNA 拷贝整合在新位点。**II 类转座因子**(转座子,transposons)它直接以 DNA 形式进行转座。通过剪切、整合或复制过程使转座子一个拷贝留在原位点,另一个拷贝整合在新位点。
- 若转座因子编码的蛋白满足它们自己转座作用(至少是转座起始,保留到细胞结束)的需要, 此类转座因子是有活性的或自主的。对于转座子而言,转座作用的最

低要求是**转座酶**(transposase)和它所识别的顺式作用位点。此位点使转座子能确定自己的DNA和宿主DNA二者之间的边界。对于反转录因子来说,在需要转座酶(一般称**整合酶**[2],integrase)时还需要反转录酶。另一些转座因子是**有缺陷的**,失去了酶的功能,但有顺式作用位点,并可以通过一个活化因子的反式作用而移动。依赖反式作用而移动的因子称**非自主的或被动的**(nonautonomous or passive)。自主的和缺陷的因子可形成相关序列的家族。

- 非自主的反转录因子的家十分庞大的且结构呈现多样性,其序列与已知的自主因子无同源性。当反转录酶和整合酶反式提供时,非自主的反转录因子作为细胞的RNA可偶然发生转座,这就产生了在脊椎动物基因组中含量很高的**加工假基因**(processed pseudogene)。
- 另一种衍生类型的移动因子,它无转座所需的顺式作用位点因此它是"稳定"的,但编码反式作用蛋白,并将蛋白给同一细胞中可顺式作用的因子,例如果蝇的P因子的**剪翅**衍生物(wings-clipped derivative)。

注释:1)**暗盒**指可交换的遗传信息的片段。大多数遗传基因座的变化发生是等位的,但暗盒是非等位的,因为发起于不同的座位。暗盒的例子包括酵母的交配型盒,锥虫的可变性表面抗原基因和细菌整合子中的抗生素抗性基因。作为此术语的一种延伸暗盒也可指用于体外诱变中可交换的寡核苷酸。2)由I类转座因子编码的**整合酶**与由II类转座因子编码的**转座酶**在结构和功能上是同源的。虽然噬菌体λ编码的位点特异性的重组酶也称为整合酶,但它不是转座酶。这两种酶在功能上是相似的,但转座酶虽无靶,但对转座因子有较高的特异性。而位点特异性的重组酶必须识别重组中供体和受体的顺序。

13.1 转座的机制

转座机制概述 根据转座机制不同,转座因子分为两个家族。**转座子**直接以DNA形式转座,而**反转录因子**以RNA为介导。每个家族中又存在几种特异类型的转座机制。在详细介绍个别机制前,先介绍适合所有转座因子的相同点。

(1) *增殖*(proliferation):所有转座事件都有在基因组中增加转座因子拷贝数的能力。此过程可被动或主动地完成。

(2) *转座的步骤*:转座子的主动转移可分为三步(见图13.1)。①**供体剪切阶段**:转座酶催化核酸内切反应,使转座因子从宿主DNA中分离出。②**链交换阶段**:转座酶催化一对转酯反应,使转座因子3′端与靶位点相连;③**修复阶段**:以DNA合成反应填补留下的缺口。对于反转录因子,它通过转录和反转录产生可加工的cDNA并进行整合来完成这三步。非病毒逆转录因子仅利用DNA修复步骤就可整合到宿主DNA中。

(3) *靶位点的重复制*:对于大多数转座因子,转座的链交换发生在靶DNA交错断裂,所以新整合的因子位于单链DNA尾端的两侧。对于缺口的修复合成产生同向重复顺序,故称为**靶位点重复**(target site duplications **TSD**)(图13.2)。大多数转座因子产生特定长度的TSDs,但I.2类的反转录因子在同一基因组中也可产生不同大小的TSDs。这可能是因为它利用偶然打开的DNA末端的原因。

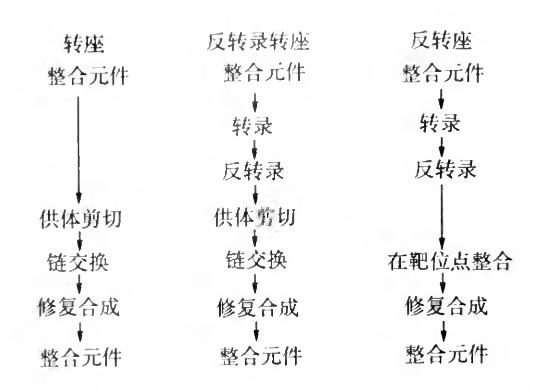


图 13.1 三种主要的自主转座机制的总结，它们分别利用 II 型，I.1 型，I.2 型转座因子。

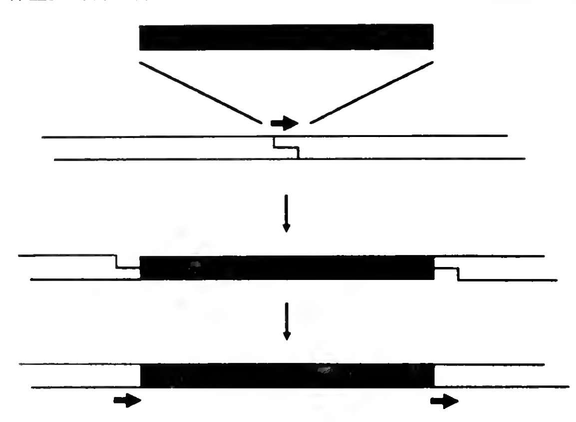


图 13.2 靶位点重复(TSD)产生的机制。靶位点重复是转座特有的标志，其他重组机制只能精确地保留重组序列，但不能使靶位点重复。在靶位点处产生交错切口，转座因子与 5′突出端连接。剩余的缺口由修复合成补平，故在交错切口产生同向重复顺序。几乎所有的转座因子均产生 TSD，不能产生 TSD 的因子(如 IS91)可能在靶位点处产生一个平末端。

(4) **靶位点的优先性**：不同转座因子对靶位点选择的优先性不同。有的在特异性位点整合，有的随机整合。大肠杆菌转座子 Tn7 一般在特异性位点整合，但若位点无效的话，它能转换为随机整合。大多数转座因子避免整合在早已存在的相同类型的转座因子上，即它们有**顺式免疫作用**(*cis*-immunity)。经过对个别转座事件的研究发现，大多数转座因子有位置优先性。一些真核转座子的移动距供体位点很近，仅(1～10kb)，这是**区域的再整合**。细菌复制型转座子优先转移到质粒上，而不是由质粒转移到细菌染色体。

保守转座和复制转座 转座机制或者是保守的或者是复制的。**保守转座**(conservative transposition，也称为**非复制转座**或**简单转座**或**切割粘贴机制**)涉及从一个位点切下转座因子而在另一位点整合。自主转座机制不增加转座因子的拷贝数，尽管可被动完成(见下文)。复制转座(replicative transposition)(又称为**非保守**或**复合转座**)。涉及转座因子的重复，一分拷贝留在供体位点，一份整合在靶位点。反转录因子转座必定是复制的，因为反转录因子产生转录本，再变成 cDNA 整合到基因组中。

对于转座子，保守转座和复制转座之间的区别反映了供体切割反应的时间不同，如图 13.3 所示。两者通过供体剪切因子的 3′端起始转座，提供可与靶位点相互作用的游离末

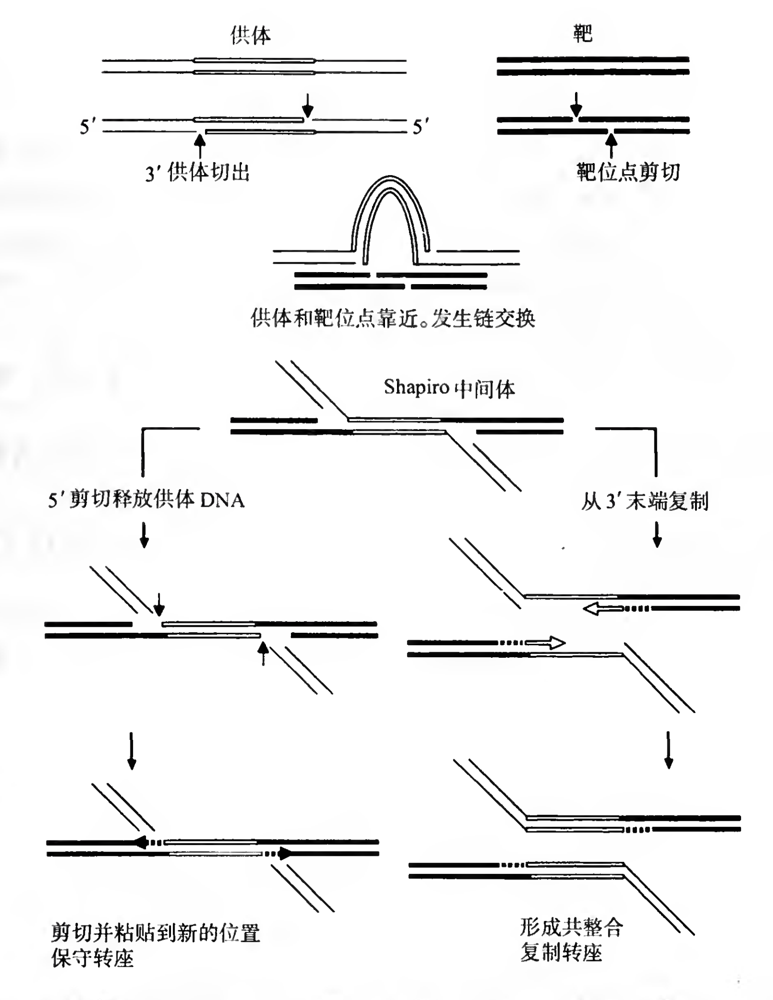


图 13.3 非复制转座和复制转座的统一模型。细线代表供体 DNA，粗线代表靶 DNA，方框代表转座子，黑色箭头代表链发生断裂的地方，白色箭头代表复制。此模型中，在形成 Shapiro 中间体前，两种转座的步骤相同。Shapiro 中间体指供体和靶 DNA 结合在一起的结构。如果供体 5′端发生切割，则供体 DNA 释放，此过程为非复制转座。如果供体 5′端不发生切割，而在连接点进行复制，此过程为复制转座。供体 5′端切割发生在链交换前，所以不形成 shapiro 中间体结构，这种机制被非复制转座子采用。若靶位点切割发生在链交换转酯反应时，靶 DNA 的双链切口不能产生(对照归巢内含子、被动转座)。

端。在保守转座中，供体 DNA 的释放是通过转座因子 5′端的切割完成的，并使供体 DNA 留下一个缺口。在复制转座中，只有转座子的 3′端被切割，所以供体 DNA 和靶 DNA 保持物理连接，修复复制以宿主 DNA 的 3′游离末端为引物越过转座子。然后转座因子复制。DNA 转座的公认模型是根据包括噬菌体 Mu、Tn5 和反转录病毒在内的转座因子的特点进行深入研究而得到的。此模型有不同的形式，把主要的反应放在不同的次序上。所有步骤中最重要的一点是供体 5′端切割的时间选择。如果它发生在链交换前，转座因子以游离分子形式释放；如果它发生在链交换后，转座因子直接从一个位点转移另一个位点。对于象噬菌体 Mu 那样既可非复制转座，又可复制转座的转座因子，其供体 5′端切割有两种可能，一般发生在链交换后。对于像 P 因子、Ac-Ds 因子那样非复制转座的转座因子，其供体 5′端转座可能发生在链交换前。目前关于 5′端切割的机制知道的很少。

解离　在两个质粒之间经复制转座后,两个环可连接出一个复合结构,称为**共合体**(cointegrate)。在共合体中,参与形成共合体的两个复制子之间的交界处是以转座因子的拷贝作为标志的。以这种方式移动的转座子也具有将共合体分离成两个复制子成分的功能,此过程称为**解离**(resolution)。每个转座子可编码**解离酶**(resolvase)。解离酶是可作用转座子内部的一段短序列的位点特异性重组酶。这段短序列称为**内部解离位点**(res,internal resolution site)。由解离酶催化共合体中两个 res 元件之间的重组将共合体解离为两个分子。解离酶的活性有很强的倾向性,即此反应很少将两个分离的分子形成共合体。这种特异性是由于拓扑学条件所导致。

被动转座　被动转座指转座因子的任何转移过程都不能依靠本身来完成。被动转座有两种类型。

(1)**反式被动转座**采用与主动转座相同的机制,但所需的酶(如转座酶、反转录酶)由主动转座因子反式提供。只有非自主的因子的转移采用反式被动转座。

(2)**细胞控制的被动转座**指由细胞 DNA 复制所导致的转座事件(图 13.4)。自主和非自主的因子采用此过程移动。

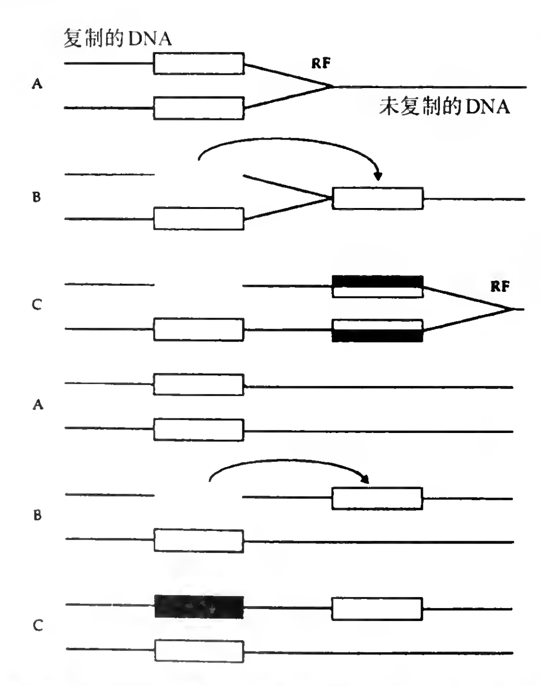


图 13.4　细胞控制的被动转座中的非复制转座使拷贝数增加。上图示意了两种机制,每种分为预转座阶段(A),主动转座阶段(B)和被动转座阶段(C)。归巢 内含子和内蛋白子采用没有主动转座的第二种机制。内含子或内蛋白子编码的核酸内切酶切割同源序列,内含子的 DNA 和序列通过基因转换拷贝到无内含子的等位基因中。

所有的转座因子必须通过转座增加自己的拷贝数以防灭绝。对于非复制转座的因子的扩增必须通过细胞被动地实现。细胞控制的被动转座以两种方式产生①从已复制的 DNA 转座到未复制的 DNA——例如玉米的 Ac-Ds 因子即采用此方式。②通过基因转变修复切除所产生的断裂——例如果蝇 P 因子,归巢内含子和归巢内蛋白子(homing intein)(参阅)采用此方式。另外拷贝数的增加可通过增加宿主复制子的拷贝数而实现,如

细菌转座子整合到质粒或噬菌体时采用此方式。

反转录转座和反转座　I.1 型的反转录因子的转移方式是通过转录出被整合因子后反转录成 cDNA，最后整合在一新位点（cDNA 合成的过程在*反转录病毒*（参阅）感染周期这一部分作更为详细的讨论）。对于反转录病毒转座作用发生在病毒感染周期的连接中，即转录本和反转录酶一起包装在病毒粒子中，然后导入细胞。而其他的反转录转座子，整个周期发生在细胞内。

一旦产生 cDNA，病毒家族的反转录因子按照转座子相似的途径转座。但因为它多了一步额外反转录，故称**反转录转座**（retrotransposition）。其 cDNA 两侧通常有很少量的宿主 DNA（一般为 2bp），这是在类似转座阶段宿主 3′端切割的反应中通过整合酶在每条链的 3′切割而产生，然后整合酶在 cDNA 的游离 3′末端和宿主靶位点间进行转酯反应。整合后是修复反应，它可产生 2bp 的靶位点重复，并可能从转座因子的 5′端除去原供体位点的 2bp 的 DNA。

I.2 型的转座因子通过与反转录病毒复制无关的机制进行转移。其 cDNA 通过以自身或靶位点上的 DNA 为引物而产生。整合过程是以 RNA 为模板，越过靶位点的 cDNA 修复合成而完成。因为此途径缺少典型转座机制中的许多步骤，但有一步涉及反转录，故称为**反转座**（retroposition）。图 13.5 表示反转录转座和反转座的机制。

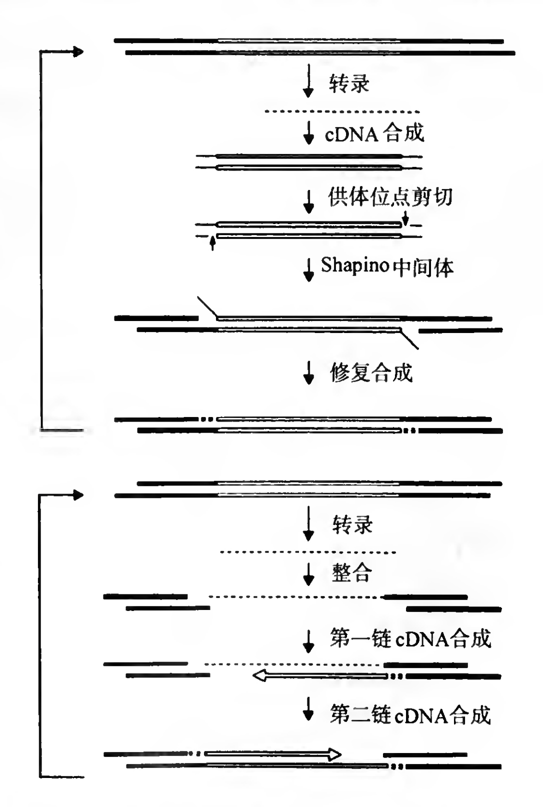


图 13.5　反转录转座机制（上）和反转座机制（下）的总结。细线代表 RNA，方框代表整合的因子，粗线代表 DNA，粗虚线代表在靶位点重复位置修复合成的 DNA。

转座的调控　作为自私DNA,成功的转座因子经常发生转移并尽可能增加它的拷贝数。但不受控制的转座会破坏宿主的基因组,并最终影响宿主的基因组中的转座因子。所以大多数转座因子必须调控它们本身的转座作用。这些调控机制不仅控制转座的频率,而且还控制转座相对细胞周期的时间和靶位点的特点。此原则的一个例外是噬菌体Mu,它的宿主大肠杆菌的存活并不重要。噬菌体Mu通过重复转座破坏宿主的基因组,然后多拷贝的噬菌体Mu与宿主DNA一起包装在噬菌体颗粒中。

转座作用有很多调控方式。果蝇P因子和细菌转座子Tn*10*是了解的最清楚的转座因子。很多转座子编码的转座酶阻遏蛋白与转座酶是共线性的。P因子在生殖细胞中转座即以此方式受到限制。P因子的转座酶基因中有三个内含子,在生殖细胞和体细胞中发生不同的拼接。在生殖细胞中,三个内含子均被除去,成熟的转录本编码一个有转座酶活性的87kDa的多肽。在体细胞中,内含子3未被除去。在此内含子中有一终止密码子,故产生截短蛋白。此蛋白可以各种方式作为转座的阻遏物——可以结合转座酶的底物,阻断其活性或结合酶本身来抑制它。相同的机制也被其他转座子利用,如玉米的Ac因子虽然它不是组织特异性的拼接。Tn*5*也可产生截短蛋白以用来选择转录起始点。

Tn*10*是通过由一个叫做p_{OUT}的启动子的转录来抑制转座,p_{OUT}和转座酶的启动子p_{IN}极性方向相反。从p_{OUT}起始的转录以各种方式阻遏转座:通过反向转录来对抗P_{IN}的产物;通过产生反义RNA以阻遏转座酶结合到识别位点。Tn*10*的转座酶是顺式优先(*cis*-preferce)的例子,即它只作用于编码它的因子上。尽管顺式优先的原理现在还不很清楚,但它代表另一类转座调控的机制:增加Tn*10*的拷贝数并不能增加有效转座酶数量,也并不增加转座活性。

转座酶的结合和从p_{IN}启动子的转录二者均可被DNA的甲基化所抑制(参阅)。Tn*10*中有许多Dam甲基化位点,所以其转座被严格地限制在复制后未被甲基化的链上的一段区域。这使得转座因子只有在宿主细胞中有第二份基因组时才能发生转座。所以由剪切造成的损伤可由重组方式得到修复。DNA的甲基化在真核生物中用以抑制转座,看来像是对高度增殖(overpriliferation)的一种拷贝数依赖的防卫系统(见第7章DNA甲基化和外遗传的调节)。

13.2　转座的效应

转座的遗传学效应　转座因子是因其可导致突变而首次被认识的。**插入突变**(基因破坏)是转座最直接的效应。但转座也可通过干扰宿主基因与其调控元件之间的关系或改变DNA的结构而影响基因的表达。

插入诱变一次能影响一个或几个基因,以致杂种不育。转座的典型作用是产生**不稳定的突变等位基因**。突变等位基因(参阅)它能以很高的频率回复为野生型状态。但回复突变不受诱变剂的影响(这是因为这种回复突变是一个剪切事件)。在细菌中,操纵子中的插入突变常有极性效应。因为转录被转座因子阻断了,或下游基因不能有效翻译(参见极性突变)。细菌的转座子通常有转录终止子或复杂的二级结构以阻止邻近启动子的转录。极性突变可防止转座因子的异位激活作用,此作用对细胞是致死的。

除了插入失活,如果转座因子中含有内源基因,并且此基因为因子内部元件调控,基

因表达的变化也可能发生。此现象常在反转录病毒中发现，并导致许多癌基因(参阅)的分离。转座因子可利用自己的调控顺序影响附近基因的表达，如反转录病毒型的转座因子的长末端重复顺序(参阅)中常有强启动子。另外，转座因子中在插入顺序上具有杂种调控元件，如许多 IS 因子中有与细菌启动子 -35 盒(参阅)基序类似的顺序，故它能激活因缺少此基序而被阻遏的操纵子，如 *lac*。

基因的表达还可因为 DNA 结构的改变而受影响。IS1 可通过改变插入处 DNA 的拓扑学而激活隐藏在内部的操纵子(相同的效应也可由编码 DNA 旋转酶的基因发生点突变而导致)。一些真核生物利用 DNA 的甲基化限制转座子的增殖，但如果转座子插入在内源重复序列附近，它们会发生甲基化作用，从而导致**外遗传的沉默**(epigenetic silencing)(见 DNA 甲基化)。基因座的沉默也可因果蝇反转录转座子 *gypsy* 因子而引起。*gypsy* 因子尽管插在距此基因数 kb 处，但仍可诱导基因的功能丢失。*gy* 因子含有作为边界元件(参阅)的重复序列，故 *gypsy* 因子标志着一段受阻遏的染色质区域。

精确转座的结构效应　转座事件可介导两侧的宿主 DNA 发生一系列的重排。最简单的情况是当供体和靶位点位于不同的分子上时。如两者位于不同的质粒上(**分子间转座**)，非复制转座使转座因子整合到靶复制子上，而在供体复制子留下一个缺口(此缺口若不被修复则可能被丢失)。复制转座产生**共合体**，即一个融合的复制子有两份转座因子。共合体可通过位点特异性地重组而解离，并产生没有发生变化的供体复制子和增加了一个插入片断的靶复制子。这些是交替进行的，如图 13.6 所示。当转座子转移到同一复制子的另一位点时，将产生新的结构区域(**分子内转座**)。在复制转座过程中，链交换时链与链间的特殊关系决定了是产生缺失还是重复倒位，如图 13.7 所示。非复制的分子间转座不产生任何重排，例如 IS 因子(尽管它在切割位点上产生一个切口)。然而复合转座子可产生大量的缺失和倒位衍生物。这些衍生物只含有一个 IS 因子，并缺少转座子的内部区域。

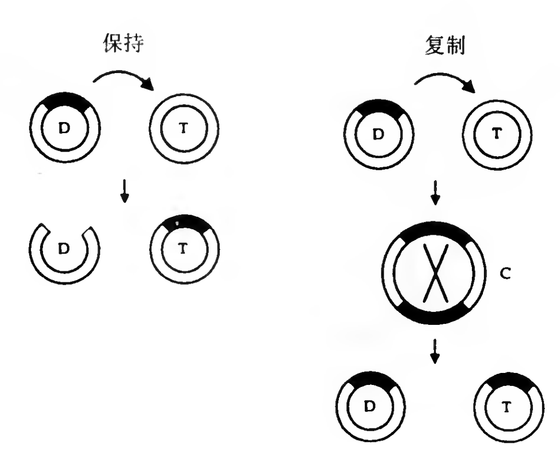


图 13.6　分子间转座的效应。D,供体分子；T,靶分子；C,共合体。交叉代表位点特异性的重组(解离)。

在非复制转座过程中，转座子的**精确剪切**不能使供体位点的结构回复到整合前状态。因为如果切口通过同向连接作用得以修复，靶位点的重复仍是转座事件的**足迹**(footprint)。根据足迹的大小和位置，经精确剪切产生的基因仍可恢复表达。如 *Spm* 因子虽

然恢复了初始的阅读框,但仍产生 3*bp* 的重复。甚至不是 3 核苷酸倍数的足迹最终也可通过第二次突变而回复。很多整合事件发生在非编码 DNA 中,此时足迹无作用。

异常转座的效应　以上所讨论的都是正常转座事件,异常或不完美的转座过程包括**一端转座**(one-ended transposition)即转座子的一端整合在靶位点;**部分转座**(partial transposition) 即只有转座子的一部分发生转座,在供体位点留下了大的足迹;**隐蔽位点转座**(cryptic-site transposition)因为转座酶识别宿主两侧 DNA,故它可与转座子一起转移;**协同转座**(cooperative transposition)两个转座子偶尔可转座间插 DNA 片段,转移的效率与两转座因子的距离成反比。细菌的复合转座子可自发完成此过程,但小的环状复制子有时转移错误的中央区域,即转移了复制子的其他区域,而不是转座子的中间部分。这种转座称为**倒位转座**(inverse transposition)。

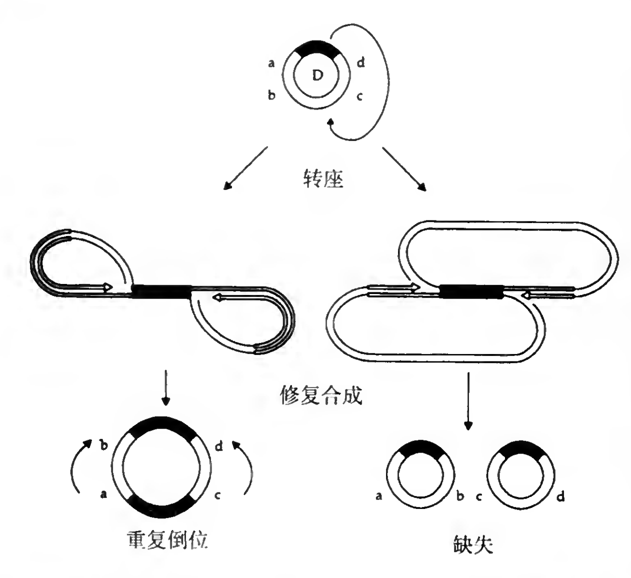


图 13.7　分子间复制转座的结构效应。供体分子(D)有 a、b、c、d 四个基因座,c 是复制原点。中间的一排表示交替的 Shapiro 中间体,当供体与靶序列在同一环状分子中时,它可形成。在一个构型中,DNA 合成穿过中间体产生反向重复。另一个 DNA 合成穿过中间体将原复制子分成两个。缺少复制起始点 c 的环将丢失,另一个环将以缺失突变体形式存在。

对宿主细胞活力的影响　宿主细胞可利用已整合因子的被动转座作为正常 DNA 代谢的一部分。但细胞 DNA 代谢可通过三个主要过程介导异常重排: ①试图修复剪切所留下的切口;②转座因子内的重组;③转座因子间的重组。

转座子的切除在供体位点产生双链缺口。细胞常试图以另一条姊妹染色单体(或细菌的姊妹基因组)或真核生物中的一条同源染色体为模板,通过同源重组进行修复。另外,断裂端可直接发生末端连接(产生简单的靶位点重复足迹),或者通过降解游离末端来加工,产生缺失。在真核生物中,由同源重组介导的修复经常导致被动转座。因为含有已整合的转座因子的等位基因可作为模板使用。如果同源染色体缺少转座子,一个没有转座子的等位基因将被用作为修复的模板。这将产生回复,即切除了转座因子和靶位点的重复。在负载等位基因的转座子中偶尔只有部分信息被用以同源重组,导致部分被动转

座，产生一种不完全切割效应。

产生双链断裂的另一反应是 **P-DNA**[1]的形成。P-DNA 是指宿主 DNA 在供体位点的反向重复。P-DNA 是复制后由姊妹染色单体在断裂的染色单体末端链到链(Strand to strand)连接形成，并产生一个发夹将两个染色单体连在一起。然后具有双着丝粒的染色体的分离可在末端产生随机断裂，在一条染色体上产生反向重复而在另一条染色体上产生缺失。新的断裂又在下一轮的复制中重新连接，并重复以上过程，因此称为**断裂-融合-桥循环**(breakage-fusion-bridge)。当端粒加在断裂端时，此过程停止(参见非法重组)。

转座因子内的重组可导致供体位点的剪切。此过程由反向末端重复顺序或同向重复顺序之间的同源重组而产生。前者常产生完全(perfect)剪切(但留下一个靶位点重复足迹)；后者常留下包含一个同向重复顺序的大足迹。后者的过程可能解释了很多真核生物中“单一”的反转录转座子 LTRs 的普遍存在，如酵母中 δ 因子，它是 Ty 反转录转座子的 LTR。靶位点重复顺序之间的重组也可发生，并导致回复。这种部分剪切、完全切割和回复都可能是转座因子内部重组的结果。在一些特殊转座因子所采用的机制中，重复序列的长度起着重要作用。

相似的过程在不依赖宿主或转座子编码的重组系统中也可发生，并留下相关的因子的足迹。此时，反向重复顺序之间可形成链间发夹或茎环结构。复制穿过环出区域时将导致子链的缺失。这是链的滑动(strand slipping)形成非法重组(参阅)的极端例子。

最后，拥有中到高度拷贝数的转座因子存在着可移动的重组序列。同一染色体上分散的同源转座因子之间的重组可导致缺失、倒位或(线性染色体的)环化。不同染色体上两个因子之间的重组可导致末端缺失、染色体的融合(形成共合体)和易位。

转座的纵览　以上讨论的转座效应都是单个转座事件的术语，但在一个特定的基因组中存在几十个到几百万个潜在的转座因子，所以有必要在更广范围来关注转座的含义。至于转座因子在基因和结构的进化和基因组形成中所起的作用，在另一章中加以讨论，(见基因组和作图、蛋白的结构功能和进化)。

果蝇中**杂交不育**指特定杂交后代中出现的许多缺失突变。不育后代常不能生育，并有很高频率的染色体重组和突变。但这些突变不稳定，可在随后的世代中回复为野生型。此现象是因为许多转座因子同时被激活而产生的。果蝇中两种转座因子家族，P 因子和 I 因子都可导致杂交不育。当 P 品系雄性(带有 P 因子)与 M 品系雌性(雌性中无 P 因子)杂交时，**发生 PM 不育**。但反交并不产生不育后代，P 品系雄性与 P 品系雌性之间的杂交也不产生不育后代。P 品系的卵细胞细胞质中含有 P 因子转座的阻遏蛋白，而 M 品系的细胞质中不含有。它们被称为 P 细胞型和 **M 细胞型**。当 P 品系的精子 DNA 释放在 M 品系的卵细胞中时，精子基因组中的许多 P 因子被同时激活，并转移到新位点，从而导致突变。相反，释放在 P 品系的卵细胞中的 P 因子会被卵细胞中的阻遏蛋白所阻遏。阻遏蛋白是 P 因子编码的一个被截短的转座酶。

在通过 DNA 甲基化全方位调节转座子移动的植物中可发现转座活性的突然爆发。许多不同家族的转座因子同时转移，这意味着大多数因子由同一种机制调控。与果蝇的

1) P-DNA 由发夹介导的修复在转座子剪切位置常产生 DNA 的反向重复。但相同的术语也可用来描述在乙醇溶剂中某些 DNA 的三级结构，还可以用于任何 DNA 在减数分裂(参阅)粗线期中的合成。

情况不同,植物转座激活时具有协同效应和再生效应,这些效应在连续发育过程中受到调控。现在不清楚此点是否反映了转座爆发在发育过程中的特殊作用。

转座因子作为一种研究工具　转座因子最显著的特点是它们可被分子生物学家作为很有用的研究工具。转座因子在基因组间转移的能力使它可作为基因传送载体使用。它们破坏基因的倾向可作为诱变剂使用。转座因子已被用于产生基因组重排,作为可转移的报告系统和克隆邻近基因的标签等,转座因子在很多方面都可被应用,请参看表 13.1。

表 13.1　转座因子的用途

用途	例子
基因传送载体	通过克隆和转化转座子在细菌染色体中导入基因
	通过向生殖细胞注射含有 P 因子的重组质粒将基因导入果蝇
	通过转移 Ti 质粒载体上重组的 Ac-Ds 因子将基因导入植物
	通过重组的反转录病毒载体产生转基因小鼠
基因突变	通过普遍性插入诱变产生随机突变体
	通过使用转座子和基因特异性引物的 PCR 分离特定的突变体
	通过剪切后的转基因介导修复产生寻靶突变
结构重组	通过分子间转座产生细菌的嵌套缺失
	通过向基因组转移 FLP/FRP 重组酶系统产生特异性的基因组重排
基因表达	通过强启动子诱导邻近基因的过量表达
	通过*截留载体*(参阅)进行内源调节功能的分析
克隆	以转座子为标签直接从突变生物体中克隆基因——*转座子标记*(参阅)
	利用*质粒拯救*(参阅)在转座子中导入质粒的复制起始点
	以噬菌体 Mu 进行体内克隆 - 通过包装到噬菌体头部从突变细胞中分离基因
基因作图	通过插入突变绘出质粒的图谱
	导入可利用测序引物的模板进行大规模测序

13.3　转座子

转座子的结构和亚分类　转座子这个术语曾描述细菌中带有抗药性基因的转座因子以及与转座相关的转座因子。现在这个术语指直接以 DNA 形式移动的所有转座因子,即所有第 II 类转座因子。

所有转座子都有一个保守结构,结构中含有一个或多个开放阅读框(其中一个编码转座酶)。两侧是**反向末端重复序列**(inverted teminal repeats,**ITRs**)。末端重复序列是转座所必需的,因为它是转座酶所识别的底物。但仅靠 ITR 还不够,还需要其他一些内部序列。

在细菌中三种主要转座子被识别。I 类转座子包括 IS 因子(简单转座子)和复合转座子;**II 类转座子**是复杂转座子;**III 类**转座子是与 Mu 噬菌体相关的转座噬菌体。有人建议将不属于以上三类的转座子称为第 IV 类。所有真核的转座子都类似细菌的 IS 因子,因为它们只编码转座所需的功能蛋白。大多数真核的转座子有典型的结构:一个中央区域,其两侧是短的 ITR。一类特殊的转座子称为折回因子(foldback elements),它主要

由很长的反向重复序列组成。对折回因子的移动结构了解很少,但它不利用 RNA 中间体。转座子结构的总结见图 13.8。

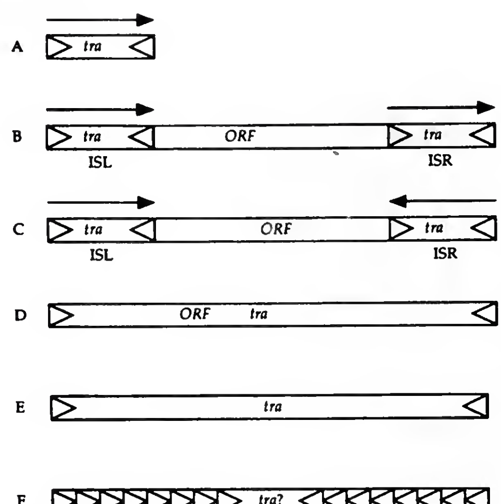


图 13.8 不同类型转座子的结构 (A) 细菌的 IS 因子;(B)含有同向 IS 因子的复合转座子;(C)含有反向 IS 因子的复合转座子。ISL 和 ISR 的名称是依据内部区域遗传图的极性而定的;(D)一个复杂转座子;(E) 一个典型的真核转座子;(F) 一个真核折回因子。*tra* 是编码转座酶的基因。在图 B 和 C 中只有一个 *tra* 是其功能必须的。*ORF* 是编码非必需蛋白的基因,如抗生素抗性。

IS 因子　细菌的转座因子的发现是因为它们能使大肠杆菌发生不稳定但是很强的极性突变。杂交分析显示一种小家族的插入 DNA 序列与观察到的很多突变有关,这些序列称为**插入顺序**(IS 因子)。大约有 100 种不同的 IS 因子已被鉴定,大多数在肠道细菌和它们的质粒中。它们一般很小(<2.5kb),由一个中央单一区域和两侧不完全的反向末端重复序列构成,这些末端重复序列大小、序列、相关性各不同。中央区域可能含有 1~3 个开放阅读框,其中一个编码转座酶。没有和转座功能无关的基因,所以它们又被称为**简单转座子**。不同 IS 在靶位点的优先性和靶位点重复顺序大小方面不同。这些特征反映转座子编码的转座酶的特异性。表 13.2 列出一些 IS 因子的属性。

表 13.2　一些 IS 因子的属性

名称	因子大小	ITR(L/R)	TSDs	优先的靶位点
IS1	768	20/23	9	以 G 或 C 为末端的 AT 丰富区
IS2	1327	32/41	5	P2 基因组的热点
IS3	1258	29/40	3	?
IS4	1426	16/18	11~13	$AAAN_{20}TTT$
IS10	1329	17/22	9	GCTNAGC
IS91	≈1800	8/9	0	?

ITRs :反向重复序列; TSDs: 靶位点重复。

注意左端(L)和右端 (R)的反向末端重复常在长度和大小不同。以上所有单位都是 bp。

复合转座子和复杂转座子 转座子原定义为细菌的转座因子，不仅带有转座所必需的基因，而且还带有转座非必需的基因，如抗生素抗性基因。那些不同于 IS 因子的转座因子称为 Tn。现在转座子的术语含盖了所有细菌的可移动因子。IS 因子被确定为简单转座子，其他因子被称为复合转座子和复杂转座子。

复合转座子或复杂转座子(composite or compound transposon) 的结构包括含有转座非必需基因的中央区域和两侧两个 IS 因子。转座酶和转座的底物由一个或两个协同转移间插 DNA 的 IS 因子提供。IS 因子可同向或反向排列，故复合转座子中有长同向重复序列或长反向重复序列(见图 13.8)。有些转座子两侧的 IS 因子相同，而另一些则不同。这种不同往往是因为其中的一个发生了一次突变，并只能提供顺式作用。复合转座子中的 IS 因子也可独自转移。转座子的维持[**转座子粘着**(transposon coherence)]反映了间插序列所产生表型的选择压力(另见*反向转座*)。

复杂转座子(complex transposon) 的结构更像单个的 IS 因子，但它包含转座必需基因和非必需基因。两种基因都位于中央区域，两侧是短的反向末端重复顺序。表 13.3 列出了复合转座子和复杂转座子中一些有代表性例子的属性。

表 13.3 细菌复合和复杂转座子的特点

名称	长度	选择标记	ITRs
复合转座子(I 类) 非必需基因的两侧是 IS 因子			
Tn*5*	5.4kb	卡那霉素抗性	长(1.5kb)。反向的 IS50，只有 IS50R 有功能
Tn*9*	2.6kb	氯霉素抗性	短(18/23bp)。同向 IS1，两个都有功能
Tn*903*	3.1kb	卡那霉素抗性	长(1.1kb)。反向 IS903，两个都有功能
复杂转座子(II 类) 必需和非必需基因的两侧是反向末端重复序列			
Tn*3*	5kb	氨苄青霉素抗性	38bp
Tn*501*	8.2kb	汞抗性	35/38bp
Tn7	14kb	三甲氧苄二氨嘧啶抗性，链霉素抗性，状观霉素抗性	≈30bp

ITRs，反向末端重复。注意 I 类转座子包括复合转座子和 IS 因子。

真核生物中的转座子家族 真核生物中，转座子被称为第 II 类转座因子(I 类因子以 RNA 为中间体移动)。这套命名体系与各类细菌转座子的命名无关。

在很多真核生物已被鉴定出转座子，虽然脊椎动物中的转座子缺乏特点而且通过基因组测序表明酵母中无转座子。大多数真核生物的转座子的结构类似于细菌转座子：一个编码转座酶和转座所需其他蛋白的中央区域，两侧是短的反向末端重复序列。折回因子是真核生物转座子中一个不同的家族，它有较长的 ITR 和一个小的中央区域，区域含有几个编码转座所需蛋白的开放阅读框。真核生物中一些主要的 II 类转座因子家族的性质列于表 13.4。

表 13.4 一些有代表性的真核转座子的属性

物种	II 类转座因子家族	ITRs	TSDs
果蝇	P	31	8
	hobo	12	8
	Mariner	28	2
	FB	大	9
线虫	Tc1	54	2
玉米	*Ac/Ds*	10/11	8
	Spm/dsom	13	3
	Mu/Mn	≈200	9
金鱼草	Tam1	13/14	3
	Tam3	12	8/5
一些真核生物	TU/Puppy	大	8

TSDs,靶位点重复。

第一个被发现的转座因子是玉米转座子,现已发现 10 个或更多不同的家族。如同其他的转座因子,它们之所以被首先发现是因为它们对内源基因有不稳定的诱变效应,故称为**控制因子**。通过对它们的遗传分析将它们列入不同的家族,并分为自主型和非自主型。分子生物学分析显示出这种二分法反映了全长有活性的因子和长度不同的缺陷因子的存在。一个例子是 *Ac-Ds* 家族失去潜在转座功能:自主的基因因子(*Ac*)长 4.5kb,编码有功能的转座酶,而解离因子(*Ds*)是可变长度缺陷因子的衍生物。Ac-Ds 家族有 10～11bp 不完全末端重复,并产生 8bp 的靶位点重复序列。相同的情况在果蝇 P 因子也存在。P 因子是所有真核生物转座子中了解得最清楚,应用最广的转座子。P 因子是在对果蝇杂种不育(参阅)现象的研究中发现的。P 因子的插入导致广泛的杂种不育。黑腹果蝇的某些品系中含有 40～50 个 P 因子,而其他品系缺少 P 因子。自主的 P 因子长 3kb,两侧有 31bp 的完全 ITRs,优先的靶位点是 GGCCAGAC,产生 8bp 的靶位点重复。在特定品系的果蝇中,2/3 的 P 因子是缺陷的,其大小范围从几百 bp 到 2.9kb。

上两例中转座酶基因呈现一个开放阅读框,并由一些内含子隔断。P 因子的剪接调节在生殖细胞和体细胞中是不同的,所以转座只发生在生殖细胞中。在 *Ac* 因子中,剪接是组成性的,所以转座并不限于生殖细胞。玉米转座的一个效应是花斑,它是不同的突变发生在不同的细胞中并被克隆繁殖的结果。P 因子和 *Ac-Ds* 因子现都已作为基因转移载体用于研究基因的表达和调控。与细菌转座子类似,这些因子也可用作诱变剂,并作为克隆基因时的标签(表 13.1)。

可转移的内含子和内蛋白子　可转移的内含子指在酵母线粒体中发现的一类转座因子。这些内含子含有一个编码位点特异性内切酶的开放阅读框。这一点与细菌的限制性内切酶很相似,但它的靶位点很长,典型的大于 20bp。这么长的靶顺序意味者它出现的频率很低,其实它在基因组中只出现一次,内含子插入在此位点中。这种内切酶不能切割宿主 DNA,除非在细胞中存在一个没有内含子的等位基因。如果这种等位基因存在,此内切酶将导致双链断裂,此断裂可被含有内含子等位基因通过同源重组而修复,这个过程导致内含子被拷贝并通过基因转变被插入到一个本来没有内含子的等位基因中(这也是被动转座的一个例子,见被动转座)。这些独特的因子称为**归巢内含子**(homing introns),

这是因为它们能回到靶基因的特定位点中。更复杂的是**归巢内蛋白子**(homing intein),这是翻译的初生肽链经过剪接产生的多肽(参阅蛋白质修饰)。它们也能编码内切酶,这些内切酶切割无内蛋白子的等位基因的开放阅读框,通过基因转变启动内蛋白子转移。

细菌的整合子　　**整合子**(integren)指含有不同结构基因组成的细菌操纵子。整合子包括调控区,编码位点特异性重组酶的开放阅读框,及重组酶的靶位点 *att I*。整合子可整合多种带有 59-碱基因子的抗生素抗性基因的盒元件。这个 59-碱基因子可被重组酶识别。它不像其他类型的盒系统(cassette system)靶位点从不空着,一个盒元件可通过非互交的重组过程取代另一个盒元件。而整合子的靶位点可以是空着的或整合一个或多个串联的盒元件。整合子可通过不同盒元件的 59-碱基因子间的重组以除去特定基因或建立稳定的连接。通过 *att I* 位点的转录促进了一个共同操纵子中所有插入基因的表达。因为像 λ 噬菌体整合子的转移是通过位点特异性重组而不是转座,所以它不是真正的转座子,称为位点特异**性转座子**。

13.4 反转录因子

反转录因子结构和分类　　**反转录因子**(retroelement)(I 类因子)指以 RNA 为介导通过反转录来移动,以 cDNA 形式整合到基因组中的遗传因子。这些因子是真核生物所特有(见反转录,如酵母和脊椎动物),并是可转移 DNA 的主要形式。反转录因子可分为两类主要的家族。**病毒家族(I.1 类转座因子)**,包括反转录病毒(参阅)和在结构和转座机制上似于反转录病毒的非病毒转座子(它们被称为**反转录转座子**)。**非病毒家族**(I.2 转座因子),包括与反转录病毒的结构和转座机制不同的自主转座子[它们被称为**反转录子**(retroposon)]因得到外来的反转录酶的反式作用而发生偶然移动的 DNA 序列。这些序列称为**反转录序列**(retrosequences),细胞 mRNA 中与之对应的是**反转录基因**(retrogenes),又称**加工假基因**(processed pseudogenes,参阅)。不同反转录因子的分类的总结见表 13.5。不同真核反转录因子的比较见图 13.9,一些例子列于表 13.6。

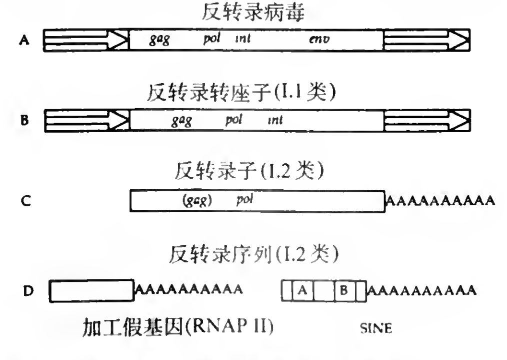


图 13.9　真核反转录因子的结构。(A)反转录病毒有 LTR(箭头)和 *gag*, *pol/int* 和 *env* 的开放阅读框;(B)反转录转座子有类似的结构但缺少 *env* 开放阅读框;(C)反转录子缺少反转录病毒结构,但有 polyA 尾,通常含 *gag* 和 *pol/int* 开放阅读框;(D)反转录序列是非自主的反转录因子,又称加工假基因。它是细胞 RNA 聚合酶 II 转录本的拷贝。它有 polyA 尾。当整合时不表达,在其他方面它们无相关性。SINE 是 RNA 聚合酶 III 转录本加工假基因。它有两个内部启动子(如 A、B 框所示),这使 SINE 能表达。

表 13.5 反转录因子的亚分类

反转录因子	定 义
反转录因子	通过反转录的 RNA 为中间体而移动的转座因子。
病毒家族(I.1 类因子)	与反转录病毒结构和转座机制相似的反转录因子。
反转录病毒	一种必须整合到宿主基因组才能复制,且复制以 RNA 为中间体的病毒。其特征是两侧具有 LTRs 和三个开放阅读框 *gag*, *pol*, *env*。其中 pol 编码反转录酶/整合酶。它只在真核生物中发现。
副反转录病毒(Pararetrovirus)	反转录病毒样颗粒,它不能转座(即因为缺少反转录酶)。它是一种缺陷反转录病毒。
反转录转座子	一种有反转录病毒特征但缺少形成感染粒子能力的移动因子(虽可形成某种胞内颗粒)。
非病毒家族(I.2 类因子)	在结构或转座机制不同于反转录病毒的反转录因子。
反转录转座子	结构不同与反转录病毒(如缺少 LTRs 和反转录病毒内部组织形式)但可编码反转录酶/整合酶的移动因子。
反转录序列(被动的反转录子)	结构与反转录子相似(无 LTR 和 polyA)但失去了反转录酶整合酶功能的移动因子,需反式提供这些功能。
反转录基因	类似内源基因的 cDNA 拷贝的反转录序列,又称加工假基因。
SINE	短散在核元件,一种起源 III 类真核基因的反转录序列(如 *Alu*,B1),它在真核生物中有成数百万个拷贝。
反转录操纵子(Retron)	细菌中未知意义的操纵子,它编码反转录酶,并有一个非翻译区。此区开始转录,然后部分反转录产生 DNA-RNA 结构。反转录操纵子不能移动,但因其序列可通过反转录而移动,故称为反转录因子。

表 13.6 真核生物中的一些反转录因子家族

物种	I.1 型因子(反转录转座子)	I.2 型因子(自主反转录子)	I.2 型因子(被动反转录子 - SINE 和其他加工过的假基因)
啤酒酵母	Ty		
果蝇	*Copia*, *gypsy*, *blood*, B104	F,I,G,*jockey*,D,*Doc*	
哺乳动物	JAP,THE1? Tdd - 1	LINE	*Alu*,B1,ID
玉米	Bs1	Cin4	

I.1 类反转录因子 反转录转座子与反转录病毒有许多结构上的相似处,包括同向长末端重复顺序,每个重复序列两侧是短反向重复序列,此因子还含有几个开放阅读框的中央区域(见有关反转录病毒基因组结构和复制循环的详细讨论)。LTR 的相似性常延伸到可识别的 U3、R 和 U5 区,相似的启动子和 poly A 位点,反转录时 tRNA 引物结合的 P 位点。反转录转座子的中央区域含有与 *gag* 和 *pol* 同源的开放阅读框。其中与 *pol* 同源的开放阅读框编码反转录酶和整合酶。但 *env* 基因通常被破坏或没有,这可解释反转录

转座子不能形成感染粒子的原因。故反转录转座子活性的机制与反转录病毒不同。反转录转座子的转录本位于核糖核蛋白体颗粒上，经过反转录整合在基因组中。通常产生2bp 的靶位点重复。这就是以上讨论过的反转录转座机制。

I.2 类反转录因子 I.2 类反转录因子代表各种不同于反转录病毒的反转录因子即非病毒家族。自主性的 I.2 类因子是反转录子。它们 缺少 LTRs 和经典的反转录病毒的组织形式，但很多都含有与反转录病毒 *gag* 和 *pol* 同源的开放阅读框。果蝇中存在反转录子的不同家族。哺乳动物中 LINE－1[长散在元件(LINE，long interspersed element)] 代表了中度重复 DNA 中重要的一类。与其他转座因子不同，I.2 转座因子产生的靶位点重复的大小不同，可能是因为偶然利用 DNA 的双链缺口。经过被动转座的细胞 mRNA 形成**加工假基因**(这些是被动反转录子)。且不能控制转座的序列。尽管加工假基因可占高等真核生物基因组的 10%，但单个基因出现的频率还是很低的。scRNA 假基因[又称**SINE**(短散在元件)]是此原则的例外。这类的特定序列是高度重复的，最好的例子是人类的 *Alu* 因子 (7SL RNA 基因的加工假基因)有将近 100 万的拷贝数。与内源基因相比因子 *Alu* 原则上已重新组织，但仍保留 RNA 聚合酶 III 内部启动子的结构，所以它是表达的。这可能与优先扩增有关。一些可移动的内含子也属于 I.2 反转座因子。这些内含子编码反转录酶，通过反转录归巢(retrohoming)而移动。

细菌的反转录酶操纵子(retron) 虽然许多细菌可合成反转录酶，转座的反转录因子只限于真核生物。反转录酶的惟一目的是产生多拷贝的单链 DNA (**msDNA**, multiple-copy，single stranded DNA)。这些 DNA 与 RNA (**msdRNA**，msDNA-associated RNA)共价连接。msDNA-RNA 的结构已在一些细菌中得到鉴定。它们的序列不同，但二级和三级结构相同。msDNA-msdRNA 的显著特征是①msDNA 的 5′端通过 2′→5′磷酸二酯键连接在 msdRNA的内部的鸟苷酸残基上；②msDNA 的 3′端和 msdRNA 的 3′端形成短杂种双链结构；③msDNA 和 msdRNA 均可形成多个茎环结构。

此 msDNA 、msdRNA 和反转录酶由单个座位编码，并被共同转录。这个座位称为**反转录酶操纵子**(retron，reverse transcriptase operqn) 。一些反转录酶操纵子是原噬菌体的一部分(这些原噬菌体为**反转录噬菌体**(retronphage))，并可作为噬菌体一部分而移动。反转录酶操纵子是不能移动的。现在还没有证据表明反转录酶操纵子的初始转录能像反转录病毒那样促进其整合到基因组中。

(张　勇 译)

第14章
诱变和DNA修复

基本概念和定义

- 突变是基因型中可遗传的变化，它由基因组中某一特定区域的核苷酸残基序列变化造成的。这些变化包括以下四种方式的**诱变**(mutagenesis)：
  (1) 在DNA复制中核苷酸的错误掺入。
  (2) 对DNA的损伤(可能是自发的，或是被环境中的诱变剂所诱导)以及细胞的修复。
  (3) 异常重组(例如，末端的连接，复制中链的跳格，不对等的交换)(参见异常重组和不对等交换)。
  (4) 可移动遗传因子的活性(见第13章可移动的遗传因子)。
- 单个碱基的错误掺入和损伤常产生**点突变**(point mutations)，也就是一个核苷酸被另一个代替(参见转换和颠换)。基因中的点突变可以改变编码蛋白的性质，那些基因外的点突变会影响一些DNA-蛋白间的相互作用。碱基的损伤还能导致致命的复制阻断。其他形式的DNA损伤(例如染色体断裂)，异常重组和可移动因子的活性往往造成更显著的**大突变**(macromutations)，这会影响到许多相邻的核苷酸。因此，突变能通过几种不同机制影响基因的功能，经常是有害影响(见第15章突变与选择)。
- 细胞将许多资源用来维持和修复DNA。它具有保证DNA复制忠实性的机制，同时有在溶液中修复被破坏的核苷酸，修复和取代DNA中被损伤的碱基并调节它们的活性以应答DNA的损伤的机制。
- 所有细胞都有一个**自发突变率**，定义为一段特定时间内基因组正常发生的突变次数。这个比例反映了DNA复制忠实性和修复的效率。控制这些过程的基因的突变可能会增高或减少自发突变率，这些基因被分别称为**增变基因**(mutator)和**抗变基因**(antimutator genes)(参见突变频率)。

14.1 诱变和复制忠实性

DNA合成过程中核苷酸的错误掺入　　尽管DNA碱基形成特定碱基对，但这不是一个高度精确的过程。从能量上考虑，预计可以产生1%～10%的出错频率。复制机构的成分通过三个主要机制将复制的精确度提高了几个数量级。

(1) DNA聚合酶的碱基选择；

(2) DNA 聚合酶的外切酶校正功能；

(3) 辅助蛋白稳定复制复合体的功能(例如单链结合蛋白,它限制了模板 DNA 二级结构的形成)。

当 dNTP 结合到模板－引物－酶的三元复合体上时,DNA 聚合酶具有高度选择性,错配核苷酸的结合没有正确配对稳定。根据特定的酶,错配的类型和其前后序列的情况的差异,鉴别错误的水平变化范围在 0～400 倍。这反映了鉴别过程中碱基配对和堆积力的影响。再有,发生在结合之后的以及将 dNTP 置于酶的活性位置的构象变化的频率,对于形成 Watson-Crick 配对的碱基来说要大 5000 倍。正确配对的碱基形成磷酸二酯键的频率也要高一些。这三种鉴别步骤的相对重要性对各种多聚酶是不同的。

校正功能由 DNA 聚合酶内在的 3′→5′外切酶活性完成。每一次加入的步骤后都有一个稍微的停顿,这让多聚酶除去错配的碱基,但是校正的主要部分发生在随后的多聚循环的起始。当引物末端和模板错配后掺入的频率非常低,这为外切酶活性提供了一个好机会。因此,校正是一种掌握时机的行为,它由多聚酶不能有效地延伸错配末端这种性质来控制。

在 *E. coli* 中,这些因素将积累的出错频率降至每 10^7 个碱基对只有一次错误掺入,但这仍比观察到的自发突变率高三至四个数量级。更高的忠实性是由**复制后错配修复**(参阅)导入的,它在复制后将新合成链中错配核苷酸纠正过来。

DNA 中瞬时的、自发的化学变化　　与 DNA 合成相联系的出错频率(在错配修复和校正之前)反映了多聚酶在选择核苷酸时的不精确性。多聚酶内在的性质(就是说,偶尔掺入错配核苷酸的趋向)是造成出错的部分原因,但这也反映了 DNA 碱基经历自发的,瞬时的变化的能力,这些变化涉及氢原子的重排。这类变化是**互变异构移位**(tautomeric shifts),而形成的同分异构体是**互变异构体**(tautomers)(图 14.1)。所有碱基以具有一定平衡的互变异构形式的混合物存在(尽管对那些出现在 DNA 和 RNA 中的碱基来说,一种形式的互变异构体很稳定,且转化成不稳定形式的时间很短)。不同形式的互变异构体可能与互补链中的不同碱基配对。例如,稳定的鸟嘌呤的酮式互变异构体只与胞嘧啶配对,但不多见的烯醇式形式与胸腺嘧啶配对。绝大多数互变异构转化是无害的,因为它们出现在双链 DNA 中且迅速回到起初状态。然而,如果在复制中模板发生一次变化,就会

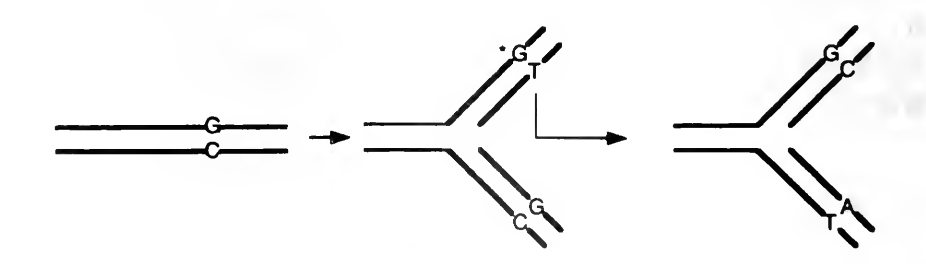


图 14.1　互变异构移位形成的诱变。鸟嘌呤正常的酮式异构体起初与胞嘧啶配对,但当作为模板时它移位成不常见的烯醇式形式(用 * 表示),造成胸腺嘧啶掺入了互补链中,这产生了错配而不是突变,因为胸腺嘧啶能被错配修复替换。如果这个胸腺嘧啶保留到下一轮复制,腺嘌呤将掺入到第三代的链中,突变就稳定下来了。

发生一次错误掺入。如果这不被纠正,一次突变将出现在下一轮复制中。起初错误掺入的碱基不一定会形成突变,因为它可能在被用作模板前被正确核苷酸取代。

移码诱变　选择和校正增加了**替换的忠实性**,也就是掺入一个核苷酸而不是另一个的正确性。DNA 聚合酶显示的另一种类型的不忠实性是**移码造成的不忠实性**,这是在模板复制时与引物链从寄存器(register)滑出时造成的。这个过程[有时称为**跳格**(slipping),**打滑**(stuttering)或**震颤**(chattering)]会产生小的插入或丢失。在编码区跳格会产生移码突变,但这个术语被用来描述所有这类性质的突变而不管它们发生在哪儿。移码诱变更依赖于模板的性质而不是 DNA 聚合酶。这类突变被重复序列(诱发跳格)和互补区域(便于形成发夹和其他二级结构)激发。这些区域是突变热点(见第 15 章突变和选择)。

移码错误会因为加工的减少而增加,这暗示着当多聚酶从模板解离后移码更易发生(这提供了一个可能的解释,为什么观察到的前导链和滞后链合成的忠实性不同,见复制)单链结合蛋白减少移码错误,它增加了单链 DNA 的稳定性且避免了二级结构的产生。

14.2　DNA 损伤:突变和致死

DNA 损伤　DNA 总是在受到攻击,遭受多种损伤。DNA 的一个损伤区域被称为一个**损害**(lesion)。DNA 损伤的影响是两方面的。①如果一个损害改变了碱基配对的专一性,它被称为**误导的损伤**(misinstructional lesion),且可能在下一轮复制中通过图 14.1 显示的机制产生一个突变。②如果一个损害使 DNA 不能特异于一个互补碱基,它被称为**非指导的损伤**(noninstructional lession),可能产生一个**复制阻断**(那儿复制又被延误了),这是致死的,因此 DNA 损伤可能产生突变或致死。为了避免复制阻断的致死效应,细胞发展了**损伤耐受机制**,这让复制能在损害存在的情况下继续(参见重组修复,SOS 反应)。产生于损害位置的突变称为**定向突变**(targeted mutation),其他位置的突变,例如归因于复制错误的突变称非定向突变(参见基因导向)。预示突变的损害称为**前诱变损伤**(premutagenic lesion)。

DNA 损伤的种类　DNA 损伤可以定义为任何不是正常完好无损 DNA 分子具有的结构,并且如果不处理的话,会造成突变和复制中断,一些类型的 DNA 损伤列在表 14.1 中。

自发和诱导损害　DNA 可能内在地(由于分子自身性质)或外在地(由于外界因素效应)遭受损伤。**自发损害**的产生由于 DNA 分子内在性质或与 DNA 损伤性化学试剂相互作用,例如作为新陈代谢副产物产生的自由基团。由这些损害产生的突变,加上那些通过复制和重组错误产生的突变,以及可移动遗传因子活动造成的突变组成了自发突变,这是特定生物自发突变率(参阅)的基础。**诱发的损害**由称为**诱变剂**(mutagens)的 DNA 损伤因子作用形成,然而特定诱变剂和天然新陈代谢副产物造成 DNA 损伤的分子机制是一样的。这儿,自发和诱变损害的区分变得模糊了。

在 DNA 中通过自发产生某些化学反应来发生损伤,这包括**脱氨**(碱基上失去氨基)和通过水解 N - 糖苷键造成**碱基丢失**,这个过程根据碱基类型称为**脱嘌呤**或**脱嘧啶**。脱氨产生误导损害,造成的后果列于表 14.2。水解造成的碱基丢失产生一个非指导性的

AP 位点(表 14.1),它阻断复制。然而在 *E. coli* 的 SOS 反应(参阅)中,转损害(translesion)的合成可能会越过 AP 位点发生,且腺嘌呤残基被优先掺入,即所谓的 **A 法则**。除了这些内在的自发反应,氧反应因子,例如自由基团,会造成一系列氧化型 DNA 损害。脱氨,碱基丢失和氧化型损伤也会被诱变剂诱导。

表 14.1 DNA 损伤的分类和它们是如何产生的

损伤类型	定 义	起 因
异常碱基 (异常碱基可以是被损伤的普通碱基或 DNA 中不适合的碱基,如尿嘧啶)	DNA 中除 A,C,G,T 外的碱基	自发脱氨 物理或化学因素造成的碱基损伤 复制过程中尿嘧啶掺入 复制过程中碱基类似物掺入 大型化学添加物加到碱基上
AP 位点 (无嘌呤或无嘧啶位点)	一个**无碱基位点**(碱基从 DNA 主干上被除去的位点)	自发碱基丢失诱发的水解 由酶将异常碱基除去产生的位点
缺口(Gap)	双链中的一条链丢失一个或多个核苷酸形成的位点	外切酶活性造成的缺口(往往与修复有联系) 由不完全复制造成的缺口(包括切除引物和复制上的旁路)
切口(Nick)	双链一条链骨架上的损伤,并没有碱基缺失。切口的特点是通过进一步破坏 DNA 末端来阻止简单的重新连接	物理/化学上的对 DNA 骨架的损伤 AP 位点的 β-去除 修复合成后的切口保存 与许多细胞过程联系的外切酶活性
断裂(双链断裂)	双链的两条链上的损害,这样分子被切开了	对 DNA 骨架的物理/化学性的损伤 与某些细胞过程联系的外切酶活性
交联	DNA 链共价连接造成的损伤,这样一个连接可能包含同一条链的两个部分(链内交联)或两条分开的链,往往是双链的两条链(链间交联)	紫外诱导的碱基二聚体化 化学诱导的交联

表 14.2 脱氨造成的结果

正常碱基	互补碱基	脱氨产物	互补碱基	后果
A	T	次黄嘌呤	C	A-G 转换
C	G	尿嘧啶	A	G-A 转换
G	C	黄嘌呤	不固定	复制中断
T	A	无		
5-甲基胞嘧啶	G	T	A	G-A 转换

四种 DNA 碱基(加上 5-甲基胞嘧啶)特异配对的碱基以及它们损伤的产物

环境中损伤 DNA 且导致突变或致死的因子是**基因毒性**(表 14.3)的。诱变剂可以是化学的、物理的或生物的因子(表 14.4)化学或物理的诱变剂通过与 DNA 作用使键断裂或重新排列,或新的键生成。这往往包含增加新的化学基团。生物诱变剂为可移动遗传因子(参阅),它们通过转座进入 DNA 而打断它,它们还造成各种类型的结构重排(见第 13 章可移动的遗传因子)。

表 14.3 基因毒性因子的分类

基因毒性剂	定　义
致癌剂	促使真核细胞肿瘤转化的因子。许多致癌物也是诱变剂,但反过来并不总是正确
断裂剂	诱导染色体断裂,也就是双链断裂的因子
诱变剂	通过直接和 DNA 作用或产生新陈代谢产物来促成诱变的因子
致癌物	诱导肿瘤形成的因子
超诱变剂	特别有效的诱变剂,例如甲基磺酸乙酯
调聚剂	导致复制提早结束的因子,例如 2′,3′-二脱氧核苷
致畸(胎)剂	诱导发育异常的因子,不一定是基因毒性因子,就是说包括不改变 DNA 结构而发挥作用的分子

表 14.4 不同种类的诱变剂

诱变剂	描　述
化学诱变剂	
脱氨剂	诱导碱基脱氨的因子,包括非专一性的硝酸和专一地使胞嘧啶脱氨的亚硫酸氢钠。
烷化剂	与核酸的亲核中心反应的分子。它用链烷以及其衍生物来取代亲核中心。这类化学性质往往是强有力的诱变剂,例如 EMS 和二甲基亚硝胺。烷化剂导致多种类型 DNA 损伤:碱基修饰会改变烷化剂的配对能力,造成复制中的错误掺入。另一方面,一个大型加合物(bulky adduct)会阻断复制。双功能的因子(能使两个亲核中心烷基化)会造成交联。
大型加合物供体	这是一种代谢后能将大型化学基团加到碱基上的分子。例如黄曲霉素 B_1 和苯丙吡。这类反应的结果是形成一个膨胀的扭曲的双螺旋,它会阻断复制。
碱基类似物	这是一些类似碱基的分子,它们能形成可以掺入正在生成的核苷酸链的核苷酸(例如,溴脱氧尿嘧啶)。与主要碱基不同的是,碱基类似物会显示异常的碱基配对性质。例如,它们能与不止一种碱基配对,因为它们在不同的互变异构形式下都能保持稳定。碱基类似物通常用一种高度专一的方式产生点突变。
插入剂	具有平面结构的分子,它能嵌入到 DNA 的碱基之间(例如溴化乙啶,丫啶橙)。插入因子通过解旋 DNA 链而增加其长度。这会诱导移码突变,阻断复制且抑制核苷酸的切除修复,这通过使酶束缚在无活性复合物状态来完成。
交联剂	这是使 DNA 链便于共价连接的分子。这些药剂因子包括双功能性的烷化剂,氮芥和硫芥,铂的衍生物。平面分子,补骨脂素当暴露于紫外光时也会交联 DNA。
物理诱变剂	
电离辐射	电离辐射导致各种 DNA 损伤,包括碱基损伤,糖环的损伤,裂口和断裂。电离辐射的效应可能是直接的(由 DNA 分子中原子电离造成)或间接的(由细胞中其他活性分子产生,主要是**活性氧组分**(reactive oxygen species),它们与 DNA 相互作用)。
紫外辐射	紫外诱导的损伤称为**光合产物**。紫外辐射的主要效应是产生相邻碱基连接的**光合二聚体**。最常见的光合二聚体是**环丁基嘧啶二聚体**,它可在任何两个相邻嘧啶间产生,T=T 是最普遍的,其他依次为 C=T,T=C,C=C.另一个常见的光合二聚体是**(6-4)损害**。包含嘌呤的光合二聚体也有形成。紫外辐射也会对单个碱基产生损伤,通常是使其水合作用。对枯草杆菌孢子的辐射会产生一个独特的**孢子光合物**。
生物诱变剂	
限制性酶	这类酶由细菌产生,它们在诸如噬菌体基因组之类的"入侵"DNA 上产生断裂(参见*限制和修饰系统*,*限制性内切酶*)。
可移动遗传因子	这是能在基因组中移动的 DNA 序列,例如病毒(比如,细菌噬菌体 Mu、反转录病毒),附加体质粒(比如 F 因子)和转座遗传因子(如 P 因子,Ty 因子)。由于它们插入或打断基因,而导致诱变的或可能带有基因调控元件来影响编码基因的表达,它们的分散的拷贝可能便于异常重组。

14.3 DNA 修复

修复机制　如果所有 DNA 损伤都不修复,细胞将很快死亡,这是因为致死突变的积累和依赖于 DNA 完整性的一些关键过程被抑制(这些过程包括复制和转录)。细胞发展了众多的机制来处理 DNA 损伤,这可被分为三大类:

(1) 直接恢复修复机制;

(2) 损伤切除和利用互补序列来修复;

(3) 可诱导的损伤耐受。

尽管有这些防卫措施,细胞最终要衰亡(例如,显然衰老的细胞)。然而,通过损伤避免和修复机制,细胞的衰弱明显地减慢了。

可被直接修复的损伤类型是很有限的。合成能特定修复每一种可能损伤的酶系所需的资源是巨大的。因此在所有有生命的有机体中的主要 **DNA 修复**方式是**切除修复**。损伤的 DNA 被除去,利用未损伤链作为模板链在缺口处合成新 DNA。这里强调了双链 DNA 基因组对许多病毒单链基因组的一个重要优势:如果一条链受损,另一条链总能被用来恢复失去的信息。直接的修复和切除因此是**准确**的,**非诱变**的或**无差错修复**(error-free repair)系统。

在以下几种情况下①双链的两条链都损伤了;②损伤出现在 DNA 的一个单链区域上(例如复制中的模板链);③切除修复系统饱和了,细胞更能承受非指导损伤而不是经受使复制中断的致死效应。在这些情况下,复制叉能穿过一处损伤,留下一个缺口,它可以通过与一同源双链重组而弥补,在相应处又留下一个原有的损伤。另一种解决方法是,一个能承受损伤的可诱导系统能发挥作用,这时 DNA 聚合酶可以用降低忠实性的复制经过损伤处。这种**倾向差错**(error-prome)的修复或**诱变修复**在 *E. coli* 中产生许多紫外诱导的突变。

DNA 修复机制对存活和保持完整性至关重要。在动物中,DNA 修复特别重要,因为影响细胞生长的突变会致癌(见第 18 章癌基因)。一些由于丢失 DNA 修复功能或失去对 DNA 损伤的细胞反应而产生的人类疾病列于表 14.5。

14.4 直接恢复性修复

DNA 损伤的直接修复　有三类损伤可以直接恢复性修复:①简单裂口能直接重连接;②某些紫外光合物能通过光复活作用来修复;③某些被烷基化的碱基能通过除去加合物修复。还有,有些酶能在受损伤核苷酸加入 DNA 前将其修复(例如 *E. coli* 中的 MutT)

裂口的直接修复　如果 3′-OH 端和 5′-P 端完好无损的话,由电离辐射等 DNA 损伤因子产生的裂口能被直接修复。然而,这类损伤往往包括对末端碱基的修饰,因此在连接以前需要进一步的加工。

表 14.5 与 DNA 修复缺陷有关的人类遗传病

病症	临床表现	遗传本质
共济失调毛细管扩张	面部血管扩张,神经肌肉退化,免疫缺陷	对电离辐射敏感显示了 DNA 修复缺陷。有关基因已经定位在人的染色体 11q 位置,尽管其功能还没确定
布卢姆氏综合征	光敏感,免疫缺陷,生长迟缓,易于癌变	高频率的姐妹染色体交换对诱变剂敏感,冈崎片段连接延迟,DNA 连接酶缺陷
科克因氏综合征(CS)	强烈的光敏感,神经错乱,矮小,视力缺陷,面部特征以及四肢尺寸不合比例	与XP类似,尽管皮肤病与皮肤癌不多见。三个互补群与三个基因被鉴定,一个是 *XPB*,其他的是 CS 专有的
范科尼氏贫血	先天性皮肤和骨骼异常,血细胞丧失	对交联剂敏感显示了 DNA 修复的职能。该病症的特征是反映染色体断裂的多重染色体突变。四个互补群与一个基因 *FACC* 被鉴定
遗传性非多发性息肉肠癌(HNPCC)	结肠直肠瘤	有关错配修复的基因的突变
毛发硫营养不良(TTD)	发脆,皮肤磷状,生长缺陷,精神迟缓,有 50%的病例光敏感	许多 TTD 细胞不能被 XP-D 细胞补体致活,这说明了 TTD 可能包含了 *XPD* 基因 不同突变用不同方式影响 XPD 蛋白,一些是影响其修复功能,其他的影响其基本转录功能(参见 TF/IF)
着色性干皮病(XP)	严重光敏感,皮癌,视力缺陷,神经错乱	七个互补群(从 XP-A 到 XP-G),另外还有 XP-V。人类基因 *XPA*, *XPB*, *XPC*, *XPD*, *XPF*, *XPG* 已被克隆。主要缺陷好象出在核苷酸切除修复(见表 14.6)

光复活　**光复活**(photoreactivation or photorestoration)是一个依靠光的 DNA 修复机制,它将某些类型的嘧啶二聚体切开(单倍体化)。这种修复途径在许多细菌中发现,例如 *E. coli* 和沙门氏菌(*Salmonella typhimurium*),但在枯草杆菌中不存在,它的主要紫外辐射产物——孢子光合物——被认为是通过一种类似但不需要光的途径单倍体化的。*E. coli* 中,用于光复合的酶有许多叫法,如 **DNA 光解酶**,**脱氧核糖二嘧啶光解酶**或**光复活酶**,它由 *phr* 基因编码。这种酶结合到具有嘧啶二聚体的 DNA 上,当暴露于波长在 300～500 nm 的可见光时,将二聚体转化为嘧啶单体。在许多但不是所有低等真核生物内都发现了性质相似的光解酶,但没什么证据证明它在高等真核生物中广泛存在[尽管在果蝇中描述了一种能直接修复(6－4)光合物的光依赖性酶]。

光复活不应与其他非酶活性的单体化机制混淆:**直接光恢复**发生在持续的对 DNA 紫外辐射下,这反映了相邻嘧啶间建立了一个单体和二聚体间的平衡。**致敏光恢复**在有色氨酸存在时发生,它提供一个电子而有助于单体化。再有,紫外辐射会扰乱其他细胞组分,抑制生长并给光合二聚体的切除修复提供时间,这种现象称为间接光修复。

烷基化碱基的直接修复　*E. coli* 中,O^6 烷基化的鸟嘌呤和 O^4 烷基化的胸腺嘧啶残

基,以及甲基化的磷酸二酯键由 **Ada 酶**修复(在其更广泛的底物专一性被发现以前,称为 O^6－甲基鸟嘌呤 DNA 甲基转移酶 I,O^6－MGT I)。Ada 具有双功能性:其一是将一个烷基基团从碱基的碱式基团上转移,另一个是将一个甲基从甲基磷酸二酯键上转移,从而直接扭转了甲基化因子的效应。Ada 被称为一个**自杀酶**,因为它将烷基转移到自身并且失去活性。第二种烷基转移酶,O^6－MGT II,不能催化甲基从磷酸三酯上转移。在其他细菌和真核生物中都找到了相似的酶,尽管它们对 Ada 蛋白不同底物的专一性不同。

适应反应　　*E.coli* 中,环境中低浓度的烷化剂能激发一个修复系统。这种适应反应是 SOS 依赖性的(参见 SOS 诱变)并和 *ada*、*aidB*、*alkA* 和 *alkB* 基因有关。此 *alkA* 基因编码 3－甲基腺嘌呤 DNA 糖基化酶,执行碱基切除修复(见下一节),识别广谱的烷基化碱基。*ada* 基因编码 O^6－MGT I,也就是刚才讨论过的 Ada 蛋白,它通过从碱基和磷酸三酯键转移烷基到自身来完成对 DNA 烷基化损伤的直接修复。从磷酸三酯键上移走的甲基被转移到一个胞嘧啶残基的 C 端。这不仅使蛋白失活,而且将其转化成其自身基因以及 *aidB*,*alkA* 和 *alkB* 的正调控因子。由该蛋白修复活动造成的蛋白失活通过诱导基因表达来补充。在 DNA 碱基烷基化频繁发生的同时,由甲基化产生磷酸三酯键只是在细胞暴露在烷化剂危险的浓度时才发生,要除去它们需要更为一致的努力。通过蛋白水解酶对甲基化失活的 Ada 蛋白的剪切可以终止适应反应,因为某些剪切产物能作为适应反应基因表达的抑制物。*aidB* 和 *alk2* 的功能不明。*ada* 和 *ogt*(编码 O^6－MGT II)基本上以低水平持续转录,但只有 Ada 蛋白是遗传调控因子,结果只有 *ada* 基因能被细胞内烷基化所诱导。

14.5　切除修复

碱基切除修复　　一种叫做 **DNA 糖基化酶**(DNA glycosylases)的酶能从 DNA 中除去特定类型损伤或不合适碱基(表 14.6)。这个过程叫**碱基切除修复(BER)**。碱基由 N－糖苷键(图 14.6)的水解除去,这种键将碱基连到 DNA 主干的糖环上。损伤的一半被切除作为游离的碱基,由此产生另一种类型的 DNA 损伤,即一个 AP 位点(参阅)。

DNA 糖基化酶有以下性质:

(1) 它们识别 DNA 中损伤的或不合适的单个特定类型碱基,或者一个相关化学加合物的小基团。

(2) 它们通过 N－糖苷键的水解除去碱基。某些糖基化酶还能通过相关的 **AP 裂解酶活性**在损伤的 3′端引入一个切口,但这个行为不是后来的修复步骤必须的,它的生理意义也不清楚。

(3) 几乎所有 DNA 糖基化酶都作用于单个碱基,对涉及多个碱基的更大更复杂的损伤没有专一性。然而,对嘧啶二聚体专一的一个 DNA 糖基化酶已在噬菌体 T4 和微球菌属中发现。

(4) 大多数 DNA 糖基化酶识别损伤或不合适的 DNA 碱基,但其中一些识别正常碱基的酶在错配修复(参阅)中也起作用。

表 14.6 DNA 糖基化酶和其底物专一性，以及 *E. coli* 中编码它们的基因和已知的功能同源的真核基因。每种酶都根据"底物 - DNA 糖基化酶(DG)"原则命名。底物通常缩写(如 uracil = Ura)

酶	物种	基因	专一性	注释
Ura-DG	大肠杆菌 啤酒酵母 人类	*Ung* *UNG* *UDG*	尿嘧啶 5-羟基尿嘧啶	
3-meA-DG	大肠杆菌	*tag* *alkA*	3-meA 3-meA, 7-meG, 3-meG, O^2-meC, O^2-meT, 5-foU, 5-羟基尿嘧啶	对亚致死剂量的烷化剂的 *alkA* 部分的适应反应(参见烷化剂的剂量)
	啤酒酵母	*MAG*	3-meA 7-meG 次黄嘌呤	
	人类	*MPG*	3-meA, 8-oxoG	
Fapy/8-oxoG-DG	大肠杆菌 啤酒酵母 人类	*fgp/mutM* *OGG1* *OGG1*	开环嘌呤(例如 fapy, 8-oxoG)	对嘌呤氧化损伤的修复
内切酶 III (thymine glycol-DG)	大肠杆菌	*nth*	嘧啶氧化的衍生物	对嘧啶氧化损伤的修复
MutY	大肠杆菌	*mutY*	A:G 中的 A 和 A:8-oxo-G 错配	错配修复
内切酶 VIII	大肠杆菌	*nei*	嘧啶氧化衍生物	
外切酶 IX				
UV-内切酶	T4 微球菌	*denV* ?	胸腺二聚体	
GT 错配	人类		GT 错配中的胸腺嘧啶	错配修复

碱基切除后的修复完成 在碱基被 DNA 糖基化酶切除后，产生了一个 AP(无嘌呤/无嘧啶)位点。这被第二类叫 **AP 内切酶**的专一性修复酶识别，它在 AP 位点 5′端引入一个切口。大肠杆菌中有两类 AP 内切酶：

(1)**外切酶 III**：起初作为具相关的磷酸酶活性的 3′→5′外切酶被鉴定(因此得名)，实际上该酶的主要生理作用是其 5′ AP 内切酶活性。由于其组成型表达，它是主要的 AP 内切酶，由 xth 基因编码。其他物种中有功能性同系物：*ExoA*[肺炎双球菌(*S. pneumoniae*)、*APE/HAP*(人类)、*APEX*(小鼠)、*BAP*(牛)和 *Rrp*1(果蝇)]。

(2) **内切酶 IV**：通常只占 AP 内切酶活性的 10%，该酶由反映氧化损伤修复重要性的超氧化物诱导。它由 *nfo* 基因编码，是啤酒酵母中发现的 *APN1* 的功能同系物。

在 AP 位点 5′端切割后，另一个酶：**脱氧核糖磷酸二酯酶(dRpase)**，对于将 5′端残基水解是必须的。由它产生一个单个核苷酸缺口。DNA 聚合酶可能从这个缺口启动修复合成，用一个核苷酸修复补钉来取代丢失的核苷酸残基。另外，损伤位置的 3′端的 DNA 可以被切除，随之是广泛的取代合成。最后，修复合成后剩下的切口被 DNA 连接酶所封闭。

核苷酸切除修复 直接恢复修复和 BER 都涉及到识别特定 DNA 损伤的酶。**核苷酸切**

除修复(NER)是一个识别广谱 DNA 损伤的系统，这些损伤包括交联，大型加合物和涉及多碱基的损伤。这些损伤被一个多功能的酶复合物除去，它产生一个被 DNA 聚合酶和 DNA 连接酶修复的缺口。现在还不清楚 NER 酶是如何识别不同损伤的：其中一些，但不是所有的，导致双螺旋变形。但天然的干扰，例如缠结，发卡结构和错配只能作为不理想的底物。损伤 DNA 通过磷酸二酯键水解被释放，作为完好无损的核苷酸，或者更常见的是作为核苷酸片段。DNA 中嘧啶二聚体的修复在可见光和黑暗情况下都发生：在可见光下，二聚体通过光复活修复(参见)，因此叫**光修复**，而**暗修复**是核苷酸切除修复，它通过另一种机制发生。

大肠杆菌中的核苷酸切除修复　　大肠杆菌中有三种主要 NER 基因：*uvrA*，*uvrB* 和 *uvrC*，它们的产物结合起来形成一个依赖 ATP 的内切酶(也叫 **UvrABC 内切酶或切除酶**，叫切除酶是因为它与切除有关。图 14.2)。两分子的 UvrA 结合到一分子的 UvrB 上，这个复合物再结合到损伤 DNA 上(UvrA 的功能可能是运载 UvrB 蛋白)。然后 UvrA 解离(其中必须 ATP 的水解)，UvrC 的结合造成一个构象变化，这让 UvrB 在损伤的 3′ 产生一个切口。然后 UvrC 蛋白产生 5′切口。UvrC 的结合和随后的**双位点切割**(bimodal incisions)需要 ATP 的结合但不需其水解。5′端的切割常在损伤处上游 8 个核苷酸的位置，而 3′端切割位点根据损伤的性质更加多变。5′和 3′的切割位置都会受前后序列影响。产生的寡核苷酸片段被 *uvrD* 基因编码的 DNA 解旋酶 II 切除。同时除去了 UvrC 蛋白，让 UvrB 横跨在两个缺口上，这可能是保护单链区域免受进一步的损伤。DNA 聚合酶 I 结合到暴露的 3′羟基上并越过缺口合成。在此过程中，同时除去 UvrB，产生一个大约 12 个核苷酸的修复补钉(patch)(这类修复有时称为**短补钉 NER**)。最后的切口被 DNA 连接酶封闭。

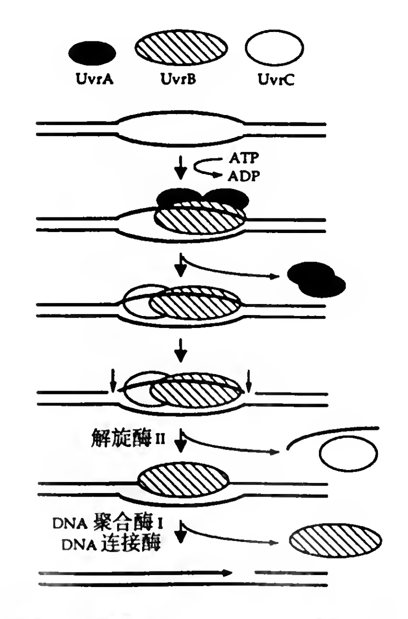


图 14.2　大肠杆菌中的核苷酸切除修复。UvrA 将 UvrB 运输到 DNA 上，然后解离；该过程中有 ATP 的水解。然后 UvrC 结合到复合物上并在损伤的 5′和 3′端引入切口(箭头处)，这时 UvrD 解旋酶将寡核苷酸片段和 UvrC 除去，让 UvrB 连接缺口。这缺口被 DNA 聚合酶 I 和连接酶修复；最后 UvrB 被去除。

UvrABC 内切酶以低水平组成型表达,但 *uvrA* 和 *uvrB* 基因也可通过 SOS 诱导(参见 SOS 反应)。第二种核苷酸切除修复过程,也与 *uvr* 基因有关且也是 SOS 诱导性的,可以造成大到 2kb DNA 的切除和取代。这种称为**长补钉 NER** 的现象,具体机制还不清楚。

核苷酸切除修补优先以活性基因已转录链为靶。当转录被 DNA 损伤中断可能会诱导切除修复。一个由 *mfd* 基因编码的称为**转录修复偶联因子(TRCF)**的蛋白能结合到拖延的 RNA 多聚酶—DNA 损伤—RNA 复合物上,并取代转录机制。TRCF还结合 UvrA,这显示一旦结合到损伤 DNA 上,它可能特异性地将 uvrA-uvrB 复合物招募到模板上。

真核生物中的核苷酸切除修复 真核生物中核苷酸的切除修复与一大群基因有关。酵母中,这些基因归到 RAD3 上位群(参见上位)。因为其中大多数已经在对辐射敏感的遗传筛选中被鉴定。人类的酵母 RAD 基因同系物起初称 ERCC(切除修复的互补成分)基因,但许多都引起特殊疾病表型,例如着色性干皮病,因而这些基因被重新命名(表 14.7)其他 NER 基因也被鉴定,因为它们对转录也是至关重要的,并显示了真核细胞转录和修复的紧密联系。这种联系的基础是普遍的转录因子 *TFIIH*,它既是细胞基本转录装置的一个重要成分,也是修复蛋白复合物**修复体**(repairosome)的主要成分。在转录复合物中,TFIIH 的核心与一个依赖细胞周期蛋白的激酶复合物(在哺乳动物中是 CDK7/细胞周期

表 14.7 与核苷酸切除修复有关的真核基因

啤酒酵母(粟酒酵母)基因	人类基因	功 能
RAD1(*rad16*)	*XPF*	RAD1 和 RAD10 结合起来形成一个内切酶,它对双链-3′单链的连接有专一性
RAD10(*swi10*)	*ERCC1*	
RAD2(*rad13*)	*XPG*	对双链-5′单链的连接有专一性的内切酶
RAD4	*XPC*	不明,可能有助于将转录复合物转变成修复体
RAD14	*XPA*	损伤趋向性 DNA 结合蛋白
RAD3(*rad15*)	*XPD*	是具有 5′→3′解旋酶活性的 TFIH 的成分
SSL2/*RAD25*(*ERCC3sp*)	*XPB*	是具有 3′→5′解旋酶活性的 TFIH 的成分
SSL1	*P44*	是基本转录因子 TFIIH 的成分
TFB1	*P62*	
TFB2	*P52*	
TFB3	*MAT1*[a]	
TFB4	*P34*	
RAD7		扰乱染色质的结构
RAD16		
RAD23		
PSO2, *PSO3*		修复补骨脂素诱导的损伤

酵母中,核心 TFII H 有七种成分,而人类核心 TFII H 有 6 种。MAT1,人类一个依赖细胞周期蛋白的激酶(CDK)装配因子,在修复中不处于核心,但它是 CDK 的一部分。酵母中的同系物 TFB3(p38)在修复时中仍留在核心中。

蛋白 H 和装配因子 MAT1,参见 CAK 结合),而在修复时,它与其他修复蛋白结合。因此,修复总是优先被引导到活性转录基因的已转录链(即**转录偶联修复**),并且转录正在进行时修复更活跃[即**转录依赖性修复**(transcription dependent repair)]。核苷酸切除修复机制在真核生物与细菌中很相似,都是先切割两端再切除之并重新合成。人类中核苷酸切除后的修复补钉比大肠杆菌中的(30nt)要稍微长一些,被含糊地称为**长补钉修复**,这用来与碱基切除修复中的 1~2 核苷酸的短补钉修复相区别。现在还没有证据证明真核生物中有像大肠杆菌中那样的大于 1500bp 的长补钉修复。

14.6 错配修复

碱基错配和错配修复 **错配修复**是一种纠正双链 DNA 中碱基对错配的切除修复过程。一个错配是一个非 Watson - Crick 碱基对,尽管错配修复系统也纠正小的(小于 4 个核苷酸的残基)插入和删除,从而不仅有助于取代的忠实性,还有助于移码突变的纠正。错配修复和其他形式的 DNA 修复不同的是被切除的物质是正常的(未损伤的)DNA 中常见的碱基,因此肯定有一种机制来区分正确和非正确的链,这样模板链中诱变的核苷酸而不是正确的核苷酸被除去。

错配能通过几种方式产生:①复制中的错误掺入;②重组时形成异源 DNA 双链;③5′甲基胞嘧啶的脱氨产生胸腺嘧啶(见 DNA 甲基化和外遗传调控)。还有,短的插入或删除能通过下面方式形成:①DNA 复制中链的跳格;②产生一个不相等的异源双链。

错配修复不影响非常相似序列间的重组,但减少了明显有差异序列间的重组,因为产生了重叠的切除区域。因此该修复系统能充当明显有差异的,但不是生殖隔离的生物体间的物种屏障。

大肠杆菌中的长补钉修复 复制中引入的错配修复被专一性地从新生链中切除,让正确序列越过产生的缺口重新生成。这称为**复制后的错配修复**,将替换和移码的忠实性提高了几个数量级。修复被引导到子链,因为它暂时还未被甲基化。GATC 顺序中的腺嘌呤残基被 DNA 腺嘌呤甲基化酶在 N^6 位置甲基化(参见 Dam 甲基化酶),但在复制后,在该酶对新合成的子单链完成甲基化之前还有一段时间差(见第 7 章 DNA 甲基化)。在这段时间里,细胞能区分亲本链和子链,因而将修复引导到子链。如果两条链都未甲基化,修复仍会出现,但不会偏向哪一边,如果两条都甲基化了,修复仍不会偏向,但效率要低得多。

四个基因(*mutH*, *mutL*, *mutS* 和 *mutD*)对长补钉修复来说是绝对需要的。它们是增变基因,因为影响它们功能的突变会改变 DNA 复制的忠实性,因而改变了自发突变率。MutS 蛋白识别错配 DNA,其效果要根据错配类型和前后序列而定(G:T 和 A:C>G:G 和 A:A>T:T,C:T 和 G:A>C:C)。未甲基化的 GATC 序列作为 MutH 的底物,MutH 是一个内切酶。子链紧靠有鸟嘌呤残基的 5′端产生切口,使包括错配核苷酸的补钉得以切除。MutS、MutL、ATP 和二价阳离子以及 MutH 的内切酶活性是需要的,并且切除时 UvrD 是必需的 DNA 解旋酶。

因为 MutH 只有在错配存在时才行使功能,所以它像是一种修复蛋白,GATC 位点和

错配必需发生物理相互作用。这种模型得到一种论证的支持。这种论证认为，无论错配位于最近的 GATC 位点的 5′端还是 3′端，切除修复总是取最短路线。从 GATC 开始而在错配的大约 100 个核苷酸以外的位置结束。由于 UvrD 是一个 5′→3′的解旋酶，这要求它处在甲基化的链来修复 5′错配，并且处在未甲基化的链来修复 3′错配，这是一个需要两个位置间相互作用的机制。根据错配的位置还需要不同方向的外切酶。图 14.3 显示了一个错配修复的模型。

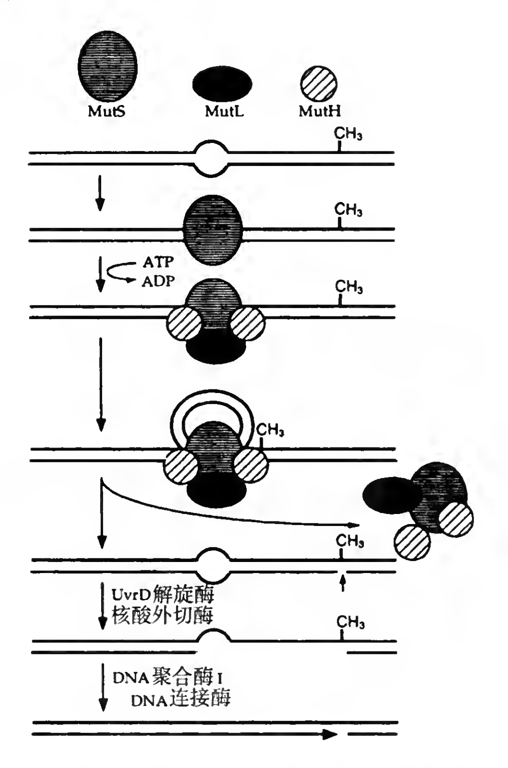


图 14.3 大肠杆菌中的一种错配修复模型。MutS 结合到一个错配位置并且募集 MutH 和 MutL。这个过程依赖于 ATP 的水解。DNA 弯曲从两侧绕过修复复合物，产生一个环，直到遇上第一个 GATC 位点。这儿 MutH 引入一个切口。UvrD 解旋酶从 GATC 位点解开 DNA，并排出修复蛋白；DNA 被一个外切酶降解。最后缺口被 DNA 聚合酶和连接酶修复。UvrD 是一个 3′→5′解旋酶，所以对于修复而言表明都必须被运载到未甲基化的链上，并使用一个 3′→5′的外切酶，例如外切酶 I。如果最近的 GATC 位点位于错配的任意一端，那么解旋酶不得不被运载到甲基化链上，并且需要一个 5′→3′的外切酶，例如外切酶 VII。

其他有机体的长补钉修复　与大肠杆菌 MutHLS 错配修复同源的系统已在其他的细菌和真核生物中被发现，尽管在许多细菌和所有真核生物中修复不依赖于甲基化。链的区分可能涉及切口的识别。这些切口是在被 DNA 连接酶修复前的复制中产生的(这些切口可能在后滞链中自然产生，而在前导链中被特异地引入，见复制)。结果，大肠杆菌 *mutS* 和 *mutL*(但没有 *mutH*)的真核同系物已被分离，列于表 14.8 中。在酵母和人类细胞中，每一种大肠杆菌基因至少有两种同系物，它们的产物形成异源二聚体。真核基因 *MLH1* 和 *MSH2* 被认为在插入/缺失错配修复中发挥主要作用，因为它们的任一个发生突变都会造成微卫星的不稳定性，这在人类中与**遗传性非多发性息肉着色性癌**(heredi-

tary nonpolyposis colorectal cancer, HNPCC)有关。

表 14.8 大肠杆菌长补钉错配修复基因的真核同系物

大肠杆菌基因	啤酒酵母基因	哺乳动物基因	注释
mutL	*MLH*1	*MLH*1	与 HNPCC 相关
	*PMS*1	*PMS*2	
		*PMS*1	
mutS	*MSH*2	*MSH*2	与 HNPCC 相关
	*MSH*6	*GTBP*	
		*MSH*1	线粒体的
	*MSH*2	*Duc*-1, *REP*3	
	*MSH*3	*MSH*3	
	*MSH*4		

短补钉错配修复 原核生物和真核生物二者都有以短切除补钉为特征的错配修复系统,特别是少于 10 个核苷酸长度的补钉。在大肠杆菌中,有两个不依赖于 *mutHLS* 错配修复的系统:

(1)**MutY 依赖的修复**取代 A:G 和 A:C 错配中的腺嘌呤残基。MutY 编码一个 DNA 糖基化酶(参见*碱基切除修复*),它的主要职能是除去腺嘌呤残基对面的 8-oxo-7,8-二氢脱氧鸟嘌呤残基,但它也能作用于错配。

(2)**极短补钉(VSP)错配修复**纠正 G:T 错配中的胸腺嘧啶。这发生在 Dcm 甲基化酶的靶顺序 CC(A/T)GG(见第 7 章 DNA 甲基化);内部的胞嘧啶残基被修饰成 5′甲基胞嘧啶并且通过脱氨可能被转化成胸腺嘧啶。VSP 修复需要 *mutS* 和 *mutL* 但不需要 *mutH* 和 *uvrD* 基因的产物。它还需要 *vsr* 编码一种内切酶,该酶对顺序 CT(A/T)GG 中的 GT 错配有专一性。

哺乳动物中一种短补钉修复已被鉴定,它纠正 G:T 错配中的胸腺嘧啶,特别是在 CpG 基序中。这同时也是一种 DNA 糖基化酶,它像大肠杆菌的 VSP 那样作用,阻止 5′甲基胞嘧啶脱氨造成的突变(见第 7 章 DNA 甲基化)。

14.7 重组修复

重组修复的证据 重组修复是涉及同源重组(参阅)在内任何的 DNA 修复或损伤耐受的战略。停留在 G_1 和 G_2 期的单倍体酵母对电离辐射的不同敏感性证实了这种修复机制的存在(这对静止期和对数期的细菌细胞也是类似的);每种情况下,具有更多基因组拷贝数的细胞抗性更强。重组修复的更直接的证据来自对大肠杆菌和啤酒酵母的重组缺陷突变体的研究,它们也是辐射敏感的。

重组修复的机制 重组修复有两种底物:双链断裂和单链缺口。对于在同源依赖性修复中有关的机制,见重组中的图 25.3。双链断裂可以通过两种方式修复,要么是一种交换修复(**双链断裂修复**),要么是一种非交换修复(**合成依赖的链退火修复**),后者出现在特

化的系统,例如P因子的转座和酵母交配型的转换。当复制在一处非指导性损伤阻断时可以产生单链缺口,并在下游重新启动合成,在损伤模板链的对面处留下一个缺口。这个缺口被一个姐妹双链的同源区域所填充[称子**链缺口修复(DSGR)**或**复制后修复(PRR)**]。利用游离的,互补的姐妹双链作为模板可对有缺口的子链进行修复合成。注意,这个过程并没有实际修复模板链起初的损伤。正如此名称所暗示的,是子链被修复了;起初的损伤仍保留在基因组中,除非通过切除去掉。DSGR是一个主要的损伤耐受机制,它在大肠杆菌中作为SOS反应(参阅)的一部分被诱导。

14.8 SOS反应和诱变修复

SOS反应的调控 大肠杆菌的**SOS反应**是对DNA损伤的一个可诱导的反应,它导致DNA修复能力增加,细胞分裂的抑制和代谢的改变。该反应由大约20个**SOS基因**(表14.9)的活化来介导,由于受到**LexA阻遏物**的转录抑制,SOS基因在正常情况下以低水平表达,该阻遏物结合在每个基因上游的操纵子序列(**SOS框**)。LexA对不同操纵子有不同的亲和性,因而SOS基因被不同的反应阈值所诱导,这些阈值导致当只有部分SOS反应被活化时才观察到的**分裂表型**(split phenotype)现象。*LexA*基因自身上游也有一个低亲和性的SOS框,因此它能抑制其自身合成,但该系统是渗漏的,此有利于对其他的SOS基因充分阻遏。当DNA被广泛损伤,单链区域被暴露出来。单链DNA与RecA相互作用

表14.9 大肠杆菌SOS基因和它们的功能

SOS基因	功能
dinB	*din* 基因是损伤诱导的基因,它们的功能还需要确认
dinD	
dinF	
dinG	
dinH	
dinI	
dnaN	DNA聚合酶III的亚基
lexA	SOS基因的阻遏物
nrdAB	DNA代谢
polB(*dinA*)	DNA聚合酶II
recA	(1) 剪切Lex A以利于SOS反应
	(2) 重组修复所必须
	(3) 剪切UmuD蛋白并且…
	(4)…在损伤处组装$UmuD_2C$复合物以利于倾向差错修复
recN	重组修复
recQ	DNA解旋酶,与重组有关
ruvAB 操纵子	重组修复
sulA	抑制细胞分裂
UmuDC 操纵子	SOS专一性的错误趋向修复
uvrA	核苷酸切除修复
UvrB	核苷酸切除修复
UvrD	DNA解旋酶,与修复有关

产生一个叫 **RecA***的活性复合物，它通过增加 LexA 本身的蛋白水解率来帮助切除 LexA 阻遏物。被切割的 LexA 蛋白不能再结合到 DNA 上，SOS 基因被解除抑制。当受损 DNA 在细胞中不再存在，RecA 失活，不再有利于 LexA 的剪切。这样细胞中高浓度的未剪切 LexA 蛋白从现存的 lexA mRNA 池中迅速积蓄，而 lexA 基因和其他 SOS 基因被关闭。这是一个典型的负反馈循环。

SOS 诱变　伴随着 SOS 反应，DNA 修复能力的增加部分由于涉及核苷酸切离和重组碱基的修复过程的组成基因的上调节（*uvrA* 和 *uvrB*——但不是 *uvrC*——具有 SOS 框，就像 *recA* 和 *recN* 那样）。在这两种机制都是精确的（**无差错修复**），同时第三种系统，其只在 SOS 反应中被诱导并涉及 *umuC*，*umuD* 基因，增强了 DNA 的修复，但也导致突变发生（**倾向差错**或**诱变修复**）。这种 **SOS 诱变**在突变前的非指导性损伤处特异性出现，也会通过忠实性降低的不可读模板链合成产生（**转损伤的合成**）。*UmuDC* 操纵子的突变，使得一些细胞对于产生复制中断的紫外辐射和其他诱变剂是不易变的，但对这些诱变剂的致死作用更加敏感，这与转损伤合成能力的丢失是相称的。这些突变体仍然对产生复制中断，导致碱基错配的诱变剂敏感。

转损伤合成的精确机制还不清楚；UmuDC 蛋白可能与 DNA 聚合酶Ⅲ相互作用并降低其选择性和（或）校对能力。依赖 UmuDC 的倾向差错修复在一般情况下被强烈阻遏，因为其活化的三个步骤都需要 RecA* 蛋白。

（1）RecA* 剪切 LexA 阻遏物来恢复 *umuDC* 的转录。

（2）RecA* 剪切 UmuD（一个无活性前体）来产生活性 UmuD′（未剪切的 UmuD 不仅在倾向差错修复中无活性，而且通过将 UmuD′束缚在无活性异源二聚体的状态来抑制倾向差错修复）。

（3）RecA* 可能有助于将 $UmuD'_2UmuC$ 蛋白复合物定位于损伤处。

SOS 反应对噬菌体感染的影响　一些病毒具有对宿主细胞指导性损伤的机制，并利用 RecA 系统。活化的 RecA* 复合物诱导剪切，不仅仅是 LexA 和 UmuD，还有噬菌体入的 CI 阻遏蛋白，因为这三种蛋白都具有一个共同的自我蛋白水解基序。CI 抑制蛋白负责维持溶原化，它的失活预示着裂解周期的到来（见第 30 章病毒，文框 30.1）。这样，通过例如紫外辐射方法激发 SOS 反应诱导 λ 原噬菌体的切离。诱导 λ 所必须的 DNA 损伤水平要比引发 SOS 反应所需要的高得多，这是因为 CI 阻遏蛋白对 RecA* 的敏感性相对低一些。基因组受损的噬菌体被引入宿主细胞后一般存活率较低。然而，如果宿主在被感染前被紫外辐射过，噬菌体能存活。这种现象，称 **Weigle 复活（W－复活）**，伴随着一个更高的噬菌体突变频率（**Weigle 诱变**）并反映了 SOS 反应的诱导导致对噬菌体基因组的迅速修复，其中一些是依赖 UmuDC 的倾向差错修复。

其他物种的趋向错误的修复　除大肠杆菌外，像 SOS 那样的倾向差错修复很少在细菌中发现，因此大多数细菌对紫外诱导的诱变有抗性，尽管更容易被杀死。功能上和 *umuC*，*umuD* 同源的基因在几种接合型质粒中发现。也有一些有限的证据证明真核生物中有诱变修复。

（张颂文 译）

第15章
突变与选择

基本概念与定义

- **突变**(mutation):在基因组的某一特定区域核苷酸序列发生改变而引起的基因型稳定的可遗传的改变[参见外突变(epimutation),副突变(paramutation)]。带有某一特定突变基因的基因组,细胞或个体被称为**突变体**(mutant)。
- 突变可能是局部的(即影响单个核苷酸或一小簇核苷酸),也可能包括基因的大片断。在前一种情况中,所谓的**基因突变**(gene mutation)发生在某一基因内部,会影响基因产物的性质或干扰其表达,而**基因外突变**(extragenic mutation) 往往不产生效应(除非它们破坏了调控元件)。大范围的突变可能涉及几十到上千个核苷酸,影响整个基因或一组基因。在真核生物中,最大的突变在细胞遗传学水平上是可见的,并被称为**染色体突变**(参见第4章染色体突变)。
- 基因突变使一个基因的等位形式变成了另一种。对许多基因座而言,**野生型等位基因**在群体中占优势。这是因为这样体现了最大的适合度(fitness)(指生存与繁殖的能力)。这种基因往往编码正常的功能性产物,此**野生型表型**(wild-type phenotype)反映了正常的基因活性。另外,少数等位基因由**突变型等位基因**(mutant alleles)所主导,那么编码的产物的量和/或结构性质可能会不同,从而产生不同的**突变体表型**(mutant phenotypes)。偏离野生型的基因突变大部分是有害的或是选择中性的,只有极少数是有益的。
- 除了单一的野生型等位基因,在一群体中可能存在着几种相同的适合度的等位基因相互平衡。那么这个基因座被称为**多态性**(polymorphic)的。如群体已处于一新发生的有益突变的过渡状态,或如存在对多个等位位点的平衡性选择,多态现象也可能涉及到不同适合度的多个等位基因。存在着多种不同的力量保持或改变一群体中等位基因的发生频率,如突变压力、迁移、随机遗传漂变及自然选择。自然选择排除了适合度降低的等位基因。所以能存活的突变的突变范围与位点具有明显的偏向,导致功能上重要的DNA序列是进化保守的。在某一环境中有害或中性的等位基因在另一环境中可能是有益的,或者有赖于它在群体中的发生频率。在现代群体某一基因座位的野生型等位基因可能在进化史的早期曾是一非常稀少的突变等位基因。
- 单细胞生物将新发生的突变传递给所有接受其染色体的后代。但在多细胞生物中,突变可能发生在生殖系统(形成的配子的组织)或躯体组织中。只有**生殖**或生

殖系统突变(germinal or gerline mutation)方可传递给后代。**体细胞突变**(somatic mutation)无法传递给后代,除非能用体细胞进行克隆。在后一情况下,根据体细胞突变发生的时间和位置不同,可能只影响单个细胞,在一野生型的背景下产生一突变细胞的克隆。或者如果体细胞突变在发育早期发生,可能产生一嵌合体(参阅)。影响生长控制基因的体细胞突变可能引发癌症(参见第18章癌基因与癌)。

15.1 突变的结构与功能性效应

突变的结构分类 根据突变涉及的DNA序列性质与数量的不同,可以分为四个结构类别。这四类突变由不同机制引起且具不同的效应。

点突变发生于单个位点并涉及少数几个核苷酸残基。如核苷酸发生在一基因内部(见表15.2),产生的效应依赖于所编码的多肽链的表达情况与结构的改变,可能是中性的,也可能具很大的危害。基因外的点突变往往是中性的,但它们如果发生在调节元件中可能改变基因表达的水平与时相。

复杂突变比较少见。它们由一簇点突变组成。一般反映了同源重组时在异源双链DNA中发生的基因转变(gene conversion,参阅),它们可能与DNA修复相联系,也可能与修复关系不大。重组序列可能是等位的(经典基因转变),也可能是非等位的(常常包括同向与反向的重复序列)。后者可能导致协同演化(converted evolution,参阅)。

大突变是染色体结构上的突变——大的缺失,插入或重排,其中最大的在细胞遗传学水平上可见。它们往往导致基因的破坏或丢失,并且有时引起基因融合效应,位置效应或基因剂量效应(见表15.3)。

染色体不平衡是染色体数量上的突变——整条染色体或偶尔整套染色体的缺失或增加。染色体不平衡引起多个基因的剂量效应,在哺乳动物中往往是致命的。

突变的功能效应——一般原理 很多突变不引起表型的变化。因为**突变位点**(mutation site,基因组内突变发生的位置)不影响基因的功能和表达或更高级别的基因功能(如DNA复制,有丝分裂分离)。这一类突变被描绘为选择性**中性**,因为它不影响个体的达尔文适合度,它们中包括很多基因外突变。很多能产生表型改变效应的突变发生在基因或能控制这些基因的调控元件内部。这种突变的靶序列有三种:基因的编码区域;位于转录单元内的非编码序列;或转录单元外的调控序列。

很多位于基因内部的点突变是中性的,这是因为它们没有改变编码产物的结构或表达情况(见表15.2)。而那些改变了其编码产物的结构或表达情况的点突变往往是有害的或是中性的——其中很少是有益的,这取决于多肽链结构的选择性约束与多肽链的功能环境(参见自然选择,分子钟)。在基因内部发生的或涉及几个基因的染色体变异往往是有害的,因为它们引起了较大的破坏(参见表15.3)。对非正常的血红蛋白的研究为很多不同类型突变的效应提供了例子(文框15.1)。

表 15.1 突变的类别与起因

点突变(point mutation)(**简单突变,单位点突变**)
定义:单个位点的突变,涉及单个残基或少量残基。**碱基替代**(base subsitutions)是点突变的一种,序列中的某一碱基被另一种碱基替换:**转换**(transitions)指一种嘌呤被另一种嘌呤所替换,或一种嘧啶被另一种嘧啶所替换;**颠换**(transversions)指一种嘌呤被一种嘧啶所替换或反之。**小的插入与缺失**(small insertions and deletions)(小于 5 nt)也可被认为是点突变。

碱基替换的起因(插入与缺失的起因见下文)
(1) DNA 聚合酶的自发错误。
(2) 复制模板含有单个碱基的误导。如碱基受损或含有不恰当的碱基,碱基类似物,或罕见的互变异构体(参阅)。
(3) 复制模板含有一单个碱基的非指导性损伤。如含有一个 AP 位点(参见 A 规则)。
(4) 5-甲基胞嘧啶脱氨基变成胸腺嘧啶。
(5) 异源双链 DNA 的错配修复。
进一步的讨论,见突变发生与 DNA 修复。

效应:依赖位点——见表 15.2。

复合突变(complex mutation)(**多点突变**)
定义:几个不连续的点突变形成的簇。

起因:
(1) 异源双链 DNA 的错配修复(或是等位基因的转变或非等位基因的,包括重复 DNA)。
(2)低忠实度的 DNA 聚合酶的活性(如反转录酶)。
(3)SOS 反应中的倾向差错的修复(参阅)。

效应:如点突变——见表 15.2。

染色体变异(macromutation)
定义:在单位点上发生,但影响很多连续的核苷酸残基。最大的大突变可以在细胞遗传学的水平上观察到。涉及 DNA 丢失的突变被称为**缺失**(deletions)与**缺损**(deficiencies)。涉及 DNA 增多的突变被称为**插入**(insertions)或**添加**(additions) [然而,如果插入的部分是基因组的成分,这种突变可称为**重复**(duplications),如果插入的部分是串联重复的一部分,那么可称为**扩增**(amplifications)或**重复扩展**(repeat expansions)]。**替代**(substitutions)是指 DNA 序列的一部分被等量的新的信息所取代的突变。**重排**(rearrangments)涉及遗传物质的重新组织,但不包括丢失与增加。重排包括**倒位**(inversions),指 DNA 的片断的方向颠倒,**易位**(translocations)指 DNA 片断从基因组的某一位置转移到另一位置;**共合**(cointegrations)或**融合**(fusions)是指两个染色体共价连接,或是线性的染色体被环化。

起因:
(1) 染色体断裂或随机的重接。
(2) 非等位基因序列的同源重组:如不等交换或不等姐妹染色单体交换,或分散的重复顺序之间的重组。
(3) 位点特异的重组。
(4) 可移动的遗传因子的活性。
(5) 串联重复序列的复制中模板或引物链发生跳格(仅仅引发缺失和插入)。
(6) 反向重复序列的复制时,模板或引物产生稳定的二级结构(仅仅引发缺失和插入)。
(7) 不平衡的减数分裂过程,涉及结构重排的染色体(仅仅引发缺失和插入)。
(8) 进一步的讨论,见第 4 章染色体突变,并参见非正常重组。

效应:依赖于突变的位点与大小——见表 15.3。

染色体不平衡(chromosome imbalance)
定义:染色体数目的异常。由于整个染色体的失去或获得(**非整数倍性**,aneuploidy)或整套染色体的失去或获得(**多倍性**,polyploidy)而引起的缺失或添加。

起因:
(1) 额外的 DNA 复制或缺乏 DNA 复制。
(2) 额外的减数分裂/有丝分裂或不发生减数分裂/有丝分裂。
(3) 染色体分离失败(往往由于事先存在的结构异常)。
(4) 涉及不平衡(非整数倍)配子的受精过程。
进一步的讨论,见染色体突变。

效应:多基因的剂量效应——见表 15.3。

表 15.2 点突变对多肽链结构与基因表达的影响

突变的位置与类别	定义及效应
位点:编码区域	
沉默突变或同义突变	定义:不影响密码子含义的碱基替换,因而对多肽链结构没有影响。沉默突变的发生是由于遗传密码的*简并性*(参阅)。例如:ATT 突变为 ATC,两者都编码异亮氨酸。 效应:同义突变一般为中性。它们在下面的情况下会影响表型,(1) 由于某一种 tRNA 非常稀少而造成的*密码子偏倚性*(参阅);(2) 突变破坏了一个内在的调控元件;(3) 突变恰好产生了一个隐秘的剪切位点。
错义突变	定义:改变密码子含义的碱基替换。导致编码的多肽链中某一个氨基酸残基被另一种氨基酸所取代。如果新的氨基酸具有与原来的氨基酸同样的化学性质,这个错义突变是**保守**(conservative)的。如果原来的氨基酸被具有不同化学性质的氨基酸所取代,错义突变是**非保守**(nonconservative)的。例如:AAG 突变为 GAG,编码的氨基酸由亮氨酸变成谷氨酸。 效应:一个错义突变的影响取决于它是保守的抑或非保守的,还取决于被替换的残基在多肽链功能中的重要程度。保守的替换一般是中性的,除非突变发生在一个关键的残基(如酶的活性位点),而非保守的替换一般会破坏多肽链的结构和/或改变多肽链的性质,从而产生突变的表型。有些错义突变的影响十分微妙,可能只在一定的环境条件下(如高温)表现出来(参见*条件突变型*)。
无义突变	定义:将一有意义的密码子改变成无义密码子(终止密码子)的碱基替换。根据突变产生的无义密码子类型可分成**琥珀突变体**(amber mutation),**乳白突变体**(opal mutation)和**赭石突变体**(ochre mutation)。例如:TGC(亮氨酸)突变为 TGA(终止密码子)。 效应:无义突变造成蛋白质合成的提前终止,产生不完整的蛋白质。突变效应的严重性由突变在编码区的位置决定。5′无义突变导致产生很不完整的多肽链,引起多肽链功能的丧失。3′无义突变可能对编码多肽链的结构影响不大,但可能影响 mRNA 的稳定性。在真核基因中的无义突变偶尔会在剪切中引起*外显子跳跃*(参阅)。
通读突变	定义:将无义密码子转变为一有意义的密码子,造成通读,使多肽链延长的碱基替换。例如:TAG(终止密码子)突变为 CAG(谷氨酰氨) 效应:连读突变可能影响多肽链的性质和 mRNA 的稳定性。一般情况下,多肽链不会延长太多,因为在天然的终止密码子的下游有不定位置的终止密码子。
移码突变	定义:一较短的 $3n+1$ 个核苷酸的插入与缺失,使阅读框发生变化,突变远端产生的新的多肽链序列。 效应:与无义突变类似。移码突变的影响取决于它的位置。5′突变的后果比 3′突变要严重。大部分开放阅读框有读框外(out-of-frame)的终止密码子,所以移码突变往往造成蛋白质合成的提前终止,产生不完整的多肽链。另见*共翻译移码*。
非移码插入/缺失	定义:一较短的 $3n$ 个核苷酸的插入与缺失。(**indel** 是插入或缺失的统称) 效应:这类突变不破坏阅读框,往往是可以容忍的。但关键残基的缺失或被隔开可能会使基因丧失功能。
位点:基因内非编码区域	
内含子	效应:内含子中的点突变一般是中性的。它们在下面的情况下会影响表型,(1) 突变破坏了一个内含的调控元件如增强子;(2) 突变破坏了一剪切位点(无法进行剪切并造成与内含子的连读,往往引起中断,有时会发生外显子跳跃);(3) 突变位点产生一个隐秘的剪切位点,使内含子的顺序包含在成熟的转录本中(一般会由于内含子中的终止密码子或是移码产生的读框外的终止密码子而造成多肽链中断)。

续表 15.2

非翻译区域	效应:很多非翻译区突变是中性的。但在有些情况下,突变可以通过影响蛋白质合成,RNA 稳定性或 RNA 转移来影响转录后调节。多聚腺嘌呤往往是有害的,因为破坏了转录后过程。
位点:基因外	
调节元件	效应:大部分基因外突变是中性的。但影响了转录调控元件如启动子或增强子的突变可能会破坏或影响转录过程。在增强子中的突变对串联基因复制(参阅)的表达图谱的多样化很重要。

表 15.3 染色体变异与染色体不平衡对基因结构,表达与功能的影响

染色体变异的物理效应	功能效应
整个基因缺失	基因功能丧失。效果取决于此位点的剂量敏感程度(参见单倍不满足性,haploinsu fficiency)。可能是一较大范围的缺失的一部分,导致一连续的基因综合症(参阅)
部分缺失(平截)	效应取决于缺失程度。较大程度的平截使基因功能丧失。不太严重的平截可能会产生具有新功能的产物。如 *v-erb* 癌基因(见第 18 章癌基因与癌)。
部分缺失(内部缺失)	效应取决于缺失程度。较大程度的缺失使基因功能丧失。较小程度的外显子的缺失可能会产生一个改变的产物,但阅读框的移动一般会导致提前终止和不稳定。在真核生物中,内含子的缺失一般是中性的,除非调控元件被改变了。
整个基因重复(基因扩增)	提高基因产物的水平。效应依赖于突变位点的剂量敏感程度。对很多基因来说,扩增没有什么影响,在一定的环境中可能还有益处(参见基因扩增)。另一些基因是剂量敏感的,因为它们的产物需要与另外的基因的产物定量作用。(参见正文中的剂量效应)
部分重复(内部)	效应取决于重复的程度与拷贝数目。外显子的重复可能会改变多肽链的功能。更小的重复(外显子内的重复)往往是有害的,因为它们改变了读框并且产生中断的不稳定的蛋白质。在一些人类疾病基因中发生基因内的串联的三核苷酸重复的延伸。(文框 15.2)
缺失/重排/插入造成的破坏	被平截的基因一般失去功能。偶尔,一个融合基因可能会具有新的性质(见正文)。如果一个基因受异源调节,那么可能存在位置效应(见正文)。
染色体丢失或获得	改变很多基因产物的剂量,往往导致严重的多剂量效应(见第 4 章染色体突变)。
整个基因组的复制	改变了所有基因的剂量。在植物中有很大的耐受性。但在较高级的动物中,由于性染色体的剂量与印迹基因的剂量在发育中起很重要的作用,这种突变的耐受性很差(参见多倍体,亲代标记)。

不论一种突变本身会引起何种效应,这种效应是否反映到表型的水平取决于另外的几种因素。

(1) 显性:在二倍体中,突变体等位基因相对于野生型等位基因是隐性的,它的影响在杂合子中不会表现出来。

(2) 遗传背景与环境:突变等位基因的效应受到非等位基因的相互作用(如丰余基因,外抑制基因的突变)或环境因素(如互养)的补偿,即使在纯合状态下也不会显示出来。

(3) 镶嵌性:在多细胞生物中,体细胞突变的效应取决于突变细胞的所占比例与分

布。

(4) 单个等位基因表达：在哺乳动物中，印迹(imprinted)基因及非活化 X－染色体上的基因(在雌性个体中)是外遗传抑制的。因此一种突变的效应取决于它的亲代来源及包含特异的非活化的 X－染色体的细胞克隆的分布(见第 7 章 DNA 甲基化及外遗传调节)。

点突变的功能性效应　大部分产生病理性效应的点突变是发生在基因编码区并且是非同义的(表 15.2)。这些突变可能通过替换单个氨基酸引起多肽链结构的特异的局部改变(错义突变)，也可能涉及更多氨基酸引起更严重的改变(无义突变，移码突变或连读突变)(图 15.1)。同义突变往往是中性的，但也可能通过偶然改变基因的表达情况来影响表型，如产生一隐秘的剪切位点。在非编码的转录序列中发生的有害的点突变往往可以用于辨明功能性重要的基序如调控元件或剪切位点。基因外侧的点突变可以通过产生或消除启动子和增强子与转录因子的结合位点或改变它们的空间关系来改变启动子和增强子的功能。

THE CAT SAW THE DOG AND RAN (END)	野生型
THE CAT SAW THE DOG AND MAN (END)	错义
THE CAT SAW THE DOG END	无义
THE CAT SAW THE DOG AND RAN AND......	通读
THE CAT TSA WTH EDO GAN DRA NEN D......	移码
THE CAT SAW THE **BIG** DOG AND RAN (END)	非移码插入/缺失

图 15.1　编码区域的点突变及其对遗传信息的解释的影响。每三个字母(核苷酸)形成一个词(密码子)，在一个句子(多肽链)中传递一定的意义。不同类型的点突变以不同方式破坏整个句子。在多肽链的开始部分发生的无义与移码突变具有最大的潜在的危害。在图中的例子中，无义突变发生在天然态的密码子附近，大部分信息的意义得以保留。反之，移码突变发生在句子的开端，大部分意义丢失了。

点突变往往只影响单个基因。但如果突变发生在处于控制地位的调控元件，如启动子，细菌操纵子(参阅)中的操纵基因位点和真核生物中的基因座控制区(参阅)，产生的后果可能会是显著的。操纵子结构基因中的突变偶尔可能通过破坏多顺反子 mRNA 的蛋白质合成来影响多个下游基因的表达(参见极性突变)。

染色体变异的功能性效应　染色体变异与染色体不平衡较之点突变效果更为严重，因为它们牵涉到更多数量的 DNA(表 15.3)。缺失与重复可能同时涉及基因的部分，整个基因与多个基因。基因部分缺失，插入(如转座因子的插入)及重复往往引起移码，从而造成序列的丢失与获得，导致非正常与不稳定的蛋白质的产生，这种蛋白质会被逐渐降解。在真核生物中，即使一分离外显子的缺失/插入也很可能导致上述效应，因为在同一基因中不同外显子的边界并不一定与读码框相关(参见外显子缺失，外显子重复)。

基因破坏(gene disruption)一般会导致功能的丧失。染色体变异与之相似，可以通过几种不同的途径来改变表型。基因融合效应(gene fusion effects)指由于缺失、重复、倒位和易位使得两个基因被打断并且使分离的元件合在一起。如果融合中保持了同一阅读框架，这个基因可以表达并产生出一个具有新的性质的杂种基因产物(例如并参见伯基特淋

巴瘤，肯尼亚血红蛋白）。缺失和重复还可能引起**剂量效应**（dosage effects），指改变细胞内基因的拷贝数而引起的病理效应。基因拷贝数改变在某些位点可能是无害的甚至是有益的（参见基因扩增），但如果细胞要求正常数量的基因产生正常数量的产物以维持与其他基因产物的竞争性平衡，那么拷贝数改变就可能引起有害的剂量效应。例如形成多聚体蛋白质要求多肽链的化学计量（如血红蛋白中的 α，β 球蛋白，见文框 15.1），或是微妙平衡的调控系统中如果蝇性别决定（参阅）的控制。另外，依赖数量调节的调控系统也是剂量敏感的——如很多细胞信号途径（见原癌基因与癌症）。染色体不平衡同时影响了很多基因的剂量。这种突变产生最严重的剂量效应（另参见性染色体非整倍性，剂量补偿机制）。最后，倒位、易位和大的缺失可能使一个基因处于非正常的调控环境中如处于一异源增强子的控制之下或进入一非活化的异染色质而失去功能。由于这种顺式作用的**位置效应**（position effect），基因表达的方式与水平可能被改变。

涉及重复 DNA 的突变　　随机的染色体断裂与末端连接发生的错误修复可能会引起缺失，插入和重排。但重复 DNA 往往归因于染色体变异。重复 DNA（参阅）可由几种途径产生[例如重复性转座，姐妹染色单体的断裂与重连接，期外复制（unscheduled replication）]，一旦发生，它可以作为非等位同源结构域，通过链跳格发生非正常同源重组及非正常 DNA 复制。

在大的同向重复单元中，如在珠蛋白基因簇中染色体可能出现错误排列，使非等位重复序列间发生同源重组。这种重组可能是同一染色体的姐妹染色单体之间的（**姐妹染色单体不等交换**，unequal sister chromatid exchange），也可能在同源染色体之间（**不等交换**，unequal crossover）。在任何一种情况下，都会导致重组一方的缺失和重组，另一方的序列插入，但不对称交换还涉及到侧翼标记的重组（图 15.2）。珠蛋白基因簇的不等交换可能引起杂种的多肽链，有时会引起致命的血红蛋白的遗传性延续（文框 15.1）。

分散的重复如转座因子之间也会发生重组。这种重组的效果取决于重组各成分的相对位置与方向。同一染色体中的同向重复之间的重组导致重复序列之间的 DNA 序列的缺失。反之，如果重复是反向的，那么重组会使间插 DNA 倒位。如果分散的重复位于不同（非同源）染色体上，重组会引起交互的易位（线状染色体）或共整合（环状染色体或质粒）。

短的串联重复序列如发生在微卫星 DNA 中的序列，在复制中往往发生链跳格，即模板与引物链配对后发生滑动，造成两条链间等同的重复序列被跳过（图 15.2）。这被认为是微卫星 DNA 的多态性形成的原因，因为在过程中不发生侧翼标记的重组（即没有交换）。链跳格复制也可由链内二级结构如发夹或十字结构引发，这些结构稳定了链的错误排列。虽然二级结构的形成受到单链 DNA 结合蛋白（参阅）的抑制从而大大提高了移码忠实度（参阅），反向重复仍是缺失与插入的热点。短的串联重复序列广泛分布在高级真核生物的基因组中，大部分是基因外 DNA，很少位于基因编码序列中。这种基因间的重复的延伸与很多人类疾病有关（文框 15.2）。

重复 DNA 序列之间的物理相互作用还导致非等位的基因转变（参阅）事件的发生。重组中间体异源双链 DNA 可能被错配修复，导致重复序列的顺序均一化。这是真核基因组中成簇点突变的主要来源（又见体细胞超突变，SOS 反应），并可能是协同演化的一种机制（参阅）。重复 DNA 序列之间的物理相互作用还导致基因表达的外遗传改变（参见

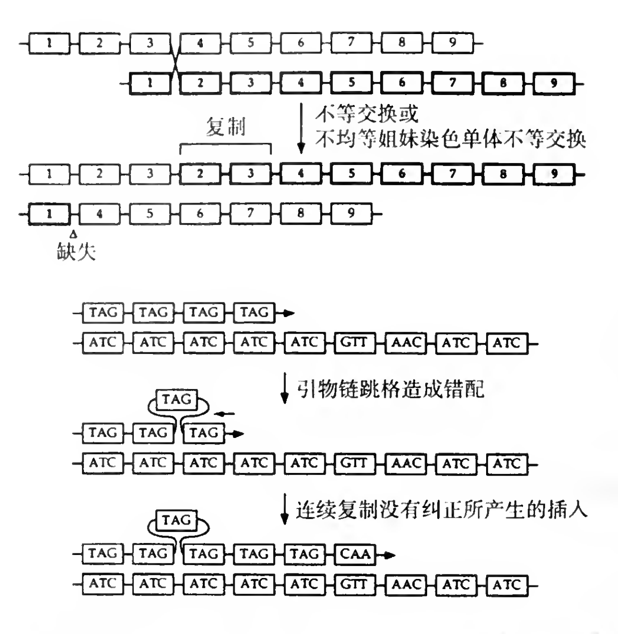


图 15.2 DNA 中通过同向重复启动插入与缺失。(A) 同向重复是不等交换(不等交换或不均等姐妹染色单体不等交换)的热点。两条姐妹染色单体可能错误排列与交换产生交互重复与缺失产物。(B) 短的同向重复是复制跳格的热点,即模板与引物链配对后发生滑动。引物链的后退跳格造成插入,前进跳格(图中未显示)造成缺失。两种跳格与不交换都与三联体重复综合征有关(文框 15.1)。

副突变,同源依赖沉默,共抑制)。

15.2 突变型等位基因及表现型的分子基础

野生型与突变型等位基因　**等位基因**(alleles)指同一基因的不同形式,开始是由基因的表型效应来定义的,但实质上还是由基因的核酸顺序决定的。任何位点的**野生型等位基因**(wild-type alleles)都是在群体中占绝对优势的等位基因,它一般具有最高的适合度,产生完全功能的产物。正向突变(forward mutation)指背离野生型的突变,它产生有别于野生型等位基因的**突变型等位基因**(mutant allele),其产物在性质、数量或分布上与野生型的产物不同。向野生型回复的突变称为**回复突变**(reversion)。

通过比较突变型等位基因与野生型等位基因的表现型,可以很简便地将其分类。在二倍体中特别重要的是在杂合子中突变型与野生型等位基因的相互作用的方式。已发展了三种分类系统来定义突变型等位基因的性质(表 15.4)。

在原则上,正向突变可以三种方式影响基因的功能与表达:它可能引起基因活性的减少或丧失(**功能丢失等位基因**,a loss of function allele);它可能使基因的活性上升,或使编码的多肽链获得新的功能(**功能获得等位基因**,a gain of function allele);或者突变型与野生型等位基因虽然核苷酸序列有差异,但不表现出表型上的差异(这种野生型与突变型等位基因被称为**同等位基因**,isoalleles)。产生同等位基因的突变是中性的,它们常常是同义替代。引起基因功能的丢失或获得的突变可能是中性的、也可能是有害的或有益的。区

分某一特定基因功能的丢失与获得的效应与整体适合度的丢失与获得是十分重要的。例如,在人类中,在某一基因座的功能的丧失导致一完全无害的不能卷舌头的表现效应(中性),而癌基因功能的获得是癌症发生途径的一步(有害)。

表 15.4 突变等位基因分类的三种系统

系统	标 准
功能的获得与丢失	可适用于单倍体与二倍体的生物体。比较野生型与突变型等位基因的功能的水平与作用范围。如突变等位基因的产物活性低于野生型,可划分为功能丢失等位基因;如产物活性高于野生型或产生新的功能,可划为功能获得等位基因。
显性关系	只适用于二倍体生物体。比较野生型,野生型与突变型的杂合子,突变纯合子的表型。根据突变的表型在杂合子中的表现程度,等位基因可分为显性,部分显性,共显性或隐性。
米勒分类	只适用于二倍体生物体。这个系统是在果蝇中发展的,果蝇具有各种可获得的缺失突变型。比较某一感兴趣的基因座缺失纯合个体与纯合突变型以及缺失/突变杂合型的表型差异。等位基因可以分为: 无效等位基因——没有活性, 亚效等位基因——相对于野生型活性下降, 反效等位基因——与野生型活性相反, 超效等位基因——与野生型相比活性上升, 新效等位基因——与野生型相比具有新的活性。

较之野生型活性下降的等位基因　　与野生型等位基因相比活性下降的等位基因被称为**功能丢失等位基因**(loss of function allele)。这可能是由于突变发生在调节基因的表达的上游区域从而减少了基因产物的量(**低表达突变**,或**启动子减效突变**,down or downpromoter mutation),也可能是突变使基因产物改变,较之野生型功能下降,例如使基因产物质量下降。如果基因表达被完全抑制,或是基因的产物完全不具有正常的功能,这样的**功能完全丧失的等位基因**(total loss of function alleles)称为**无效等位基因**(null alleles, amorphs)。整个基因的缺失、破坏调控元件的突变或使编码的多肽链提前终止的点突变都可能产生无效等位基因。如果突变只是减弱基因的功能而非完全消除,这些基因编码的生物活性在生物体仍能作用,只是相对于野生型效率降低。这种**功能部分丧失的等位基因**(partial loss of function alleles)称为**渗漏等位基因**(leaky alleles)或**亚效等位基因**(hypomorphs)。基因内部发生错义碱基替换,产物仍具有较小活性或是基因上游调控因子发生突变,使基因的表达减少但未完全消除。这两种情况都可能产生渗漏等位基因。突变表型的严重程度取决于突变多肽链的残余功能:产生一较剧烈的表型的等位基因被称为**强等位基因**(strong alleles),而产生一较温和的表型的等位基因被称为**弱等位基因**(weak alleles)或**温和等位基因**(moderate alleles)。一缺失突变的严重程度在不同环境中也可能不同(参见条件突变型)并且在环境不恒定的情况下表现出不完全的外显率。

在二倍体中,功能丢失的突变等位基因相对于野生型往往是隐性的(参阅),即突变效应在杂合子中不表现。这是因为在大部分基因座中,基因的一个野生型拷贝(它产物的

50%的剂量)就足以满足细胞的需要。

在两种情况下,功能丢失的突变等位基因相对于野生型可能是显性的。一种情况是发生**单倍不满足性**(haploinsufficiency)即维持野生的表型需要基因的两份功能性拷贝,50%的产物剂量对于基因的生理功能是不够的。在具有单倍不满足性的基因座,功能丢失的突变等位基因相对于野生型是部分显性(参阅)的(突变的效应在产物剂量59%的杂合子中表现,但在产物剂量为0的纯合子中更为严重)。一个单倍不满足性的例子是高胆固醇血症,一种低密度脂蛋白受体水平50%的减少引起的部分显性的人类疾病。

如果突变产物干扰野生型产物的功能,功能丢失的突变等位基因相对于野生型等位基因也可能是完全显性的。这常常发生在基因产物是多聚体的情况下,突变型可与野生型多肽链相互作用形成无活性的复合物。例如,受体酪氨酸激酶是二聚体,突变的不能进行信号传导的受体可以通过与野生型多肽链形成无活性的异二聚体来阻断信号。这种性质的等位基因被称为**显性失活**(dominant negative)(有时被分类为**反效等位基因**,antimorphs,因为它们与野生型作用相反或相拮抗)。

较之野生型活性提高的等位基因　相对于野生型活性有所提高的等位基因被称为**功能获得等位基因**(gain of function allele)或**超效等位基因**(hyper-morph)。这种等位基因可能通过以下途径提高基因产物的活性:提高基因产物的量(**高表达突变**或**启动子增效突变**,up or uppromoter mutation),编码产物相对于野生型具有改善的或新的性质,使基因产物的表达与活化超出野生型的常规水平(例如**组成型突变型**,constitutive mutant,野生型产物受调控表达,而突变型一直活化,**异位表达突变型**,ectopic expression mutant,调节的突变型使一个基因的表达超出常规的时间与空间区域)。如果基因的表型是明显的,功能获得突变等位基因相对于野生型往往是显性的,但也有可能在形成寡聚蛋白情况下,野生型多肽链可遮蔽功能提高突变等位基因的量的效应。**显性激活**(dominant positive)与显性失活相反,即突变型多肽链通过在寡聚体蛋白质中与野生型产物相互作用来发挥其效应,且在这种情况下突变型突破了野生型产物所受到的限制与调节,例如在组成型信号受体酪氨酸激酶中观察到的那样。**新效等位基因**(neomorph)较之野生型等位基因具有新的活性。异位表达突变型往往具有新的效应,因为在其常规的时间与空间之外表达一多肽链的效果是很难预料的。功能获得的同源异型突变型,如果蝇触角足,在触角应该发生的体节生出足,就是新效等位基因(参见*同源异型基因*)。

与野生型活性相似的等位基因　很多突变没有表型效应,并且是选择性中性的。不同等位基因具有相同的表型,无法从形态上区分彼此,这样的等位基因被称为**同等位基因**(isoalleles)。同等位基因由不改变基因产物的量、性质与分布的突变产生(例如同义核苷酸替代),或者突变虽可以改变编码多肽链的结构与表达但在表型上没有反映,因为这种改变可以忽略(例如调节区域的突变使基因表达的速率发生温和的但没有症状的改变,或在功能不重要的多肽区域发生的保守错义突变)。

同等位基因是选择中性的,但并非所有中性的等位基因都是同等位基因。如果突变引起表型的变化,但不同表型不影响适合度,突变仍是中性的(例如引起眼睛颜色改变的突变)。突变的效应可从几个层次进行考虑:① 对核苷酸序列的影响;② 对基因活性的影响;③ 对表型的影响;④ 对整体适合度的影响。只有影响到适合度的突变要经受自然

选择。对基因活性没有影响的突变(如同义突变或发生在编码区以外的大部分突变)在所有的环境中往往都是中性的。改变了基因的活性或表型的突变可能在某些环境中是中性的而在另一些环境中不是。

中性的等位基因不经受自然选择。因而可能有几个等位基因以相对较高的频率在一个群体中保持平衡(参见*多形现象*)。表型有区别的中性等位基因可从形态上区分,但同等位基因只能从分子水平区分。在某些情况下,**蛋白质多态现象**(protein polymorphisms)可从蛋白质的不同形式在电泳胶上的不同行为来监测(另见*蛋白质平截测试*)。**DNA 序列的多态现象**(DNA sequence polymorphisms)有时可通过限制酶切片断或 PCR 产物长度的改变来监测,这是因为突变可能引起酶切位点的产生或破坏(**限制性片断长度多态性**,restriction fragment length polymorphisms),或是卫星 DNA 串联重复序列的延伸或缩短(**简单序列长度多态性**,simple sequence length polymorphisms)。另外,单链 DNA 或异源双链 DNA 可用于监测突变(参见*突变筛选*)。唯一能毫不含糊地监测与分辨所有多形性的方法是 DNA 序列分析。据估算,人类 DNA 的**平均杂合性**(mean heterozygosity)是 0.004,即每 300 个碱基中有一个是具有多态性的。

15.3 突变的分布与分子进化

突变谱与突变在基因组中的区域分布 在基因组的群体中突变的谱型与分布不是随机的,有**突变热点**(mutation hotpots)的存在。突变热点指特别容易受到某种 DNA 损伤或重排的位点。串联重复 DNA 是链跳格复制或不等交换的热点,反向重复是二级结构诱导的缺失的热点,而 5-甲基胞嘧啶残基通过脱氨基是 C→T 替代的热点。5-甲基胞嘧啶的不稳定性部分解释了在哺乳动物中出乎意料的转换的频率高于颠换。对任何一个碱基,颠换有两种选择而转换只有一种选择,所以随机的改变造成的颠换的频率应是转换的两倍。事实正好相反,转换的频率约是颠换的两倍。不同碱基的替代频率相差很大。由于 5-甲基胞嘧啶的不稳定性,C→T 的转换频率是其他碱基替换的十倍,至少在具有甲基化 DNA 的生物体内。在其他 11 种甲基替代中也有一定的趋向性,这反映了 DNA 修复机制中的偏向性,特别是错配修复或由于甲基损坏造成的常见非结构性损坏的修复(见第 14 章诱变与 DNA 修复)。

当考虑到突变热点与碱基修复的偏向性,原则上,基因组的任何区域具有相同发生突变的机会,即突变是随机与非定向的,作为被作用物为自然选择提供"选择随机性"。在文框 15.3 中我们将讨论所谓的**程序性突变**(programmed mutation),即突变作为发育程序的一部分由细胞诱导发生,以及**定向突变**(directed mutation),即对特定选择压力做出的反应。

在群体中等位基因的发生频率 在一指定的群体中等位基因的发生频率受很多因素的影响,如群体的大小与结构,交配方式,突变速率,迁移,随机遗传漂变以及基因转化(见表 15.5)。突变与迁移是使新的等位基因进入一个群体的两个因素。新产生(或到达)的等位基因相对于现有的野生型等位基因可能是有害的,中性的或(极少情况下)有益的。在最简单的例子中,使适合度下降的等位基因会被自然选择消除(**负选择**,negative selection)并且由于频发突变与迁移保持一较低的频率(**突变压力**,mutation pressure)。使适合

度提高的等位基因将在群体中传播开来(**正选择**,positive selection),并逐渐代替原来的野生型等位基因。如果在这个替代的过程中对此群体进行分析,就会发现在这个位点上的**多态性**(polymorphism)(即指同时有两个或两个以上的等位基因,每一个的发生频率都大于0.01)。这可被称为**过渡性多态现象**(transient polymorphism)因为所有的等位基因的发生频率在趋向**固定**(fixation,频率为0或1)。

表15.5 描绘等位基因在群体中分布与改变分布的作用力的术语

术语	定义
突变速率	在一段时间内发生的突变的数量。例每一代每一个基因发生的突变数。
突变频率(等位基因频率)	在一指定的群体中带有特定的突变型等位基因的个体的数量。
突变压力	通过频发突变提高特定的等位基因的频率的作用力。
突变负荷(遗传负荷)	在群体中有害的等位基因的效应。
迁移	通过向群体中输入或从群体中输出个体来改变特定等位基因的发生频率的作用力。
自然选择	通过消除在一指定环境中使适合度下降的等位基因来改变等位基因的频率的作用力。
随机遗传漂变	通过随机取样改变等位基因频率的作用力。

如果一个新的基因是选择中性的,等位基因的发生频率更多地取决于概率事件,即配子的随机取样(**随机遗传漂变**,random genetic drift),而非自然选择。漂变的等位基因发生频率渐趋固定,但在一个大的群体中这需要很长的时间,到达固定后显示出**中性多态性**(neutral polymorphism)。漂变的效果在小的群体、群体的瓶颈及新的群体中较明显(建立者效应),在这些群体中出现特定等位基因时漂变可引起快速与显著的变化。

如果杂合子的适合度落在由两种纯合子的适合度定义的范围之外,情况将更为复杂。当一个等位基因较之群体中的其他等位基因适合度更大,而杂合子的适合度介于两种纯合子之间,那么选择是定向的并导致最适等位基因的固定。如果杂合子的适合度大于任何一种纯合子(超显性),选择趋向杂合子,两种等位基因将维持一种**平衡多态性**(balanced polymorphism)。这方面的一个例子是β珠蛋白的正常等位基因(Hba)与镰状细胞等位基因(Hbs)的超显性选择。在一些非洲国家具有这两种等位基因的多态性,因为虽然Hbs在纯合状态下是有害的,但HbA/HbS的杂合子具有疟疾抗性,较HbA纯合子的适合度更高。另一种平衡选择的形式包括具有频率依赖的适合度的等位基因。杂合子的适合度可能低于任何一种纯合子(次显性)。当这种情况发生时,就会出现发散的选择,但如果随机的交配持续,杂合群体将增多且各种等位基因维持一种**不稳定的多态性**(unstable polymorphism)。

选择压力与分子钟 当考虑到突变分布的区域性差别,整个基因组的DNA发生突变的可能性是相等的。然而,发生在非编码区域的绝大部分突变是中性的,而发生在编码DNA的很多突变是有害的因而被自然选择消除。因而我们观察到的现存的在编码区的突变频率远远小于非编码区,导致编码DNA序列在进化时间内是保守的。

维持了DNA编码序列的**选择压力**(selection pressure)控制了观察到的突变的谱带与分布。在编码DNA中的大部分现存突变是碱基替换,因为更剧烈的突变,如移码突变,大的缺失等,一般都是有害的并将被消除。在**简并性位点**(degenerate sites,替换不会改变密码子的含义的位点,一般在密码子的第三位)比**非简并位点**(nondegenerate site,突变产生错义,导致氨基酸替换的位点)碱基替换发生较频繁。进化中的蛋白质中非同义的碱基替换的速率表征了维持蛋白质结构的选择压力的强度。一些蛋白质改变很少因为几乎所有的残基在维持蛋白质的结构和功能中都起着重要的作用(例如组蛋白)。另一些蛋白质进化很快,因为其多肽链结构对功能不太重要(如胰岛素连接链,其功能是分开A链与B链,并在多肽链剪切后被弃去)。

同义替代的速率与选择压力无关,因为它们一般是中性的。因而,组蛋白与白蛋白的同义突变的频率是相似的,虽然在白蛋白中非同义突变的频率上百倍地高于组蛋白中的频率。这引出了**分子钟**(molecular clock)的概念,以中性进化的速率来衡量进化时间。然而,分子钟并非恒定的。不同基因中性替代速率不同即氨基酸取代速率不同,如胰岛素的中性替代的速率是次黄嘌呤转磷酸核糖基酶(HPRT)的两倍,虽然它们的非同义替代的速率相似,即处于相似的选择压力下。很多因素会影响替代的速率,其一就是DNA修复的偏向性。在上面讨论过,某些类型错配与碱基损伤较其他类型更易得到修复,所以中性进化的速率可能受基因的碱基组成的影响。基因组之间修复效率不同也会有影响(这个原因可解释动物线粒体基因组的快速进化,因为它缺少核苷酸切除修复,参阅)。分子钟在不同的种属谱系中运行速度也不同。一般,具有较长的代间隔的种属中分子钟较慢,因为新产生的突变的效果将在延续的子代中得到测试。真实时间中的代数越少,可测试的新的突变就越少。

15.4 遗传分析中的突变

经典遗传分析和反向遗传学　剖析生物体系的经典方法是分离此体系中突变缺陷型,然后决定突变基因的结构与精确的功能。另一种途径则包括基于某一基因的结构分离基因——一般通过对其编码的蛋白质的序列测定——然后突变这个基因来研究其功能,这种方法有时被称为**反向遗传学**(reverse genetics)(参见体外突变,基因导向,基因敲除)。经典的遗传分析要求突变型的产生,对群体进行筛选来辨别感兴趣体系的突变型,然后对效应基因进行作图与克隆。一旦得到感兴趣的突变型,就开始第二轮筛选来得到改变第一轮突变表型的突变,这样可以辨别那些与在第一轮辨别的基因相互作用的基因(文框15.4)。

遗传筛选　突变可以通过将生物群体暴露于物理或化学的诱变因素而产生。亚硝基胍常常用于诱变细菌类,甲基磺酸乙酯或X射线常常用于果蝇。另外,转座子可用于产生突变:P因子被广泛地用于果蝇中,*Ac-Ds*因子用于植物中(参见P因子突变发生,转座子标记)。用这种方法可产生随机突变个体的群体,然后必须分辨与分离感兴趣的体系的突变型。

在很多情况下,突变型具有特定的形态学上的表型,可以通过费力的可见筛选来得到(如大规模筛选植物、果蝇、斑马鱼的发育突变型,酵母的细胞周期突变型,果蝇与小鼠种

系的**增强子捕获**与**基因捕获**(参阅)的有趣的表达图式)。当找到一生化或生理的突变,以利通过选择得到所需突变型群体是有价值的。对于功能获得突变型,如细菌类对抗生素抗性的获得,可以使用**正选择**(positive selection or direct selection)——即在有抗生素存在下培养细菌,没有抗性的细胞(非突变细胞)被杀死。功能丢失突变,如细菌中的**营养缺陷型**(参阅)可以使用**负选择**(negative selection,或**反选择**,counter selection)。这个策略利用突变型细胞失去的某些敏感性。在营养缺陷型的情况下,反选择往往通过**青霉素富集**(penicillin enrichment)来进行:青霉素能阻止细胞壁成分的合成,因此对扩增细胞是致死的。突变型由于其代谢缺陷不能在基本培养基上扩增,因而对青霉素具有抗性(另见**正－负选择**)。青霉素富集不能分辨特定的代谢缺陷型,选择需要通过一间接的两步过程进行**影印铺板**(replica plating)。这包括将一有细菌克隆的主盘复印到织物上(将一绒布铺在克隆上,使每一克隆的部分细胞转到布上),再将这织物放到一新鲜的平板上,这样细胞沉积于培养基上,产生了同原来一样的克隆排列图式。如主盘提供了适宜的代谢终产物,原养型与营养缺陷型都可以生长,而复制平板包含基本培养基,营养缺陷型无法生长。那么就可以通过在复制盘上不出现对应克隆在主平板上辨别营养缺陷型(另见**重组选择**)。

同样的遗传筛选的原理可应用于更高级的生物,但只有在能饲养与突变大量生物的情况下才可行,例如果蝇,酵母,植物,培养的动物细胞。由于二倍性这种筛选变得更复杂,为了辨别隐性突变,突变型需要通过传代获得突变纯合子或者在一单倍的背景下进行研究(如使用非整数倍细胞系)。在有些二倍体种属中,单倍体也能生存(例如很多植物,斑马鱼)。一旦辨别了一种突变,就可以分离与克隆这个基因并了解其生化作用。这方面工作的原理另外讨论(见第 24 章重组 DNA)。

条件突变型　对于必需体系如 DNA 复制、发育、细胞周期等的遗传分析十分困难,因为突变往往是致命的。不能进行复制的细胞会死亡,在发育早期被阻断的机体也会死亡。在二倍体中,隐性的致死突变可在杂合子中保持,其致死的趋势可通过使杂合子相交,分析其纯合后代来研究。其实,不论是单倍体还是二倍体,都广泛利用**条件突变型**(conditional mutant)来分析必需体系。条件突变型带有只在特定的**限制条件**(restrictive conditions)下明显表现出来的(往往是错义的)突变。在通常的**允许条件**(permissive conditions)下,它表现出野生型的表型。条件突变型的较重要的类别有**温度敏感突变型**(temperature-sensitive mutations),提高温度条件下表现突变型表型,以及**冷敏感突变型**(cold-sensitive mutations),在低温下表现突变型的表型。在两种情况下突变型的性质都涉及在限制温度蛋白质的变性趋势上升。

遗传途径分析　很多基因是**遗传途径**(genetic pathways)和**遗传网络**(genetic networks)的部分。遗传途径指将一个底物转变为产物经过多个中间阶段,每一个阶段受不同基因控制。在某些情况下,底物指一种信息,如一个到达细胞表面的信号必须传导到核内,或是以基因调控的形式,通过多个调控开关到达下游的靶序列。在另一些情况下,底物是一个物理分子代谢物必须转变成有用的产物。突变分析可以决定这些途径中包含的基因,它们的活性顺序,以及在途径的何处发生了分支与收敛。

在某些方面代谢途径是最易剖分的,因为其起始底物,中间物以及终产物都是物理分子而非信息的传递状态。一般情况下,代谢包括一序列化学反应,每一反应都由特异的酶

催化。每一种酶都由一个基因编码。破坏酶功能的突变将造成**代谢阻断**(metabolic block)其特征是:①不能产生反应的终产物;②代谢中间物的积累。这两种不正常状态之一或共同产生表型,并且往往是对机体有害的。细菌可以利用简单的碳源,水,金属合成很多必须的有机分子,因而可以在包含这些物质的**基本培养基**(minimal medium)上生长。具有这类野生型的代谢性质的细菌细胞被称为**原养型**(prototrophic)。所谓的**营养缺陷型**是(auxotroph)指一种细菌的代谢酶的突变缺陷型,这种缺陷使营养缺陷型的细菌不能在可以满足野生型细菌的培养基上生长,而必需包含被破坏的代谢途径的终产物的**补加培养基**(supplemented medium)。如果代谢阻断的表型主要表现为不能产生代谢终产物,那么这个途径中任何基因的突变都会产生类似的表型(**基因座不均一性**,locus heterogeneity)。开始时,途径中的阶段的数目可以通过*互补分析*(参阅)来估算,将两种突变放在一起,观察彼此基因组的产物是否可以补偿另一个的缺陷。基因作用的顺序可以通过分析代谢中间物来建立。可用互养来了解是否一个突变细胞的中间产物可被另一突变型细胞用来产生终产物。

信息转移途径的分析较为复杂,因为突变既可以产生在代谢失常中最常见的功能丢失效应,也可能引起功能获得效应(组成途径活化)。显性的功能获得突变有一定用处,两个突变杂交获得同时携带这两种突变的个体,可用于研究途径的顺序。早期作用的功能丢失的突变可以阻断信号转移,它对于作用于较晚期的功能获得的,引起组成型信号转移的突变来说是下位性的。而较晚作用的功能丢失的突变相对于较早作用的功能获得的突变是上位性的(参见*上位性*,*下位性*)。

文框 15.1 人类疾病中的突变与病理学——血红蛋白失常

正常与不正常的血红蛋白　　**血红蛋白**(hemoglobin)是红细胞中的携氧蛋白,使红细胞可以从呼吸系统中运输氧。血红蛋白是一个四聚体蛋白,包括两个 α 型珠蛋白链以及两个 β 型珠蛋白链,它们通过氢键相互作用。每一个珠蛋白链与一个血红素结合,其功能是与氧结合。人类珠蛋白基因被发现有两个基因簇。**α-珠蛋白基因簇**包含 ζ-珠蛋白基因,两个等同的 α-珠蛋白基因,一个未知功能 θ-珠蛋白基因。**β-珠蛋白基因簇**包含 ζ-珠蛋白基因,两个 γ-珠蛋白基因,其产物相差一个氨基酸残基,还有 δ-珠蛋白基因,β-珠蛋白基因。这两个簇都含有假基因。

这两个簇的基因都按时间顺序表达,所以在发育过程中合成的血红蛋白在改变(β-珠蛋白基因簇的发育调控的分子基础将在文框 29.3 中讨论)。在生命的前 6~8 星期,血红蛋白在卵黄囊中合成,是由两个 ζ-珠蛋白链与两个 ε-珠蛋白链组成的四体(**胚胎血红蛋白**,embryonic hemoglobin or Hb Gower I)。然而,从第二星期开始,胚胎血红蛋白的合成下降并且 α-珠蛋白与 γ-珠蛋白的合成起始。到出生时,血红蛋白主要由肝合成,是由两个 α-珠蛋白链与两个 γ-珠蛋白链组成的四体(**胎儿血红蛋白**,fetal hemoglobin,**HbF**)。在成体中,α-珠蛋白继续合成,但在怀孕 30 星期和出生 12 星期后,γ-珠蛋白的合成下降而 β-珠蛋白与 δ-珠蛋白的合成上升。血红蛋白产生的首要位点从肝转移到骨髓。**成体血红蛋白**(adult hemoglobin)主要是 HbA($\alpha_2\beta_2$),HbA2($\alpha_2\delta_2$)占整体的一小部分(约 2%)。

血红蛋白的失常有三种形式：**血红蛋白病**(hamoglobinpathies)(珠蛋白链的结构性质改变，导致产生了不正常的血红蛋白多肽链)；**地中海贫血**(thalassemias)(珠蛋白合成的量的减少，使α-珠蛋白链与β-珠蛋白链之间出现不平衡)；发育失常(破坏了珠蛋白表达的发育时间相)。这些失常，效果从无症状到致命各不相同，阐明了很多不同类型的突变的病理效果。

点突变产生的变异珠蛋白　很多不同的珠蛋白变异体产生于错义突变，在α-珠蛋白链与β-珠蛋白链中超过50%的残基都被发现有替代。很多替代在对血红蛋白功能的影响上是中性的，而有一些是病理性的，因为干扰了血红蛋白的三级结构或进行构象改变的能力，从而改变了分子对氧的亲和力或干扰了分子与血红素的结合。最常见的病理性替代是将β-珠蛋白链的第6位密码子从GAG改变为GTG(谷氨酸被VAL取代)。产生的血红蛋白(HbS)在脱氧状态下更易发生分子间的黏附。HbS在低氧张力下结晶，形成僵硬的镰刀状的红细胞，会阻断毛细管床，损伤内脏。在突变的纯合个体中，这种细胞的破坏导致了严重的**镰细胞贫血**(sickle-cell anemia)。HbS的等位基因是很稀少的，但在某些非洲国家存在多态性，因为正常基因与HbS基因的杂合子对更严重的疟疾有抗性(参见*超显性选择*)。虽然杂合子携带有HbS基因，但很少显示出症状——只有在氧气张力特别低的情况下(**镰状细胞性状**，sickle-cell trait)。将收集的细胞暴露于极端的条件下，可以识别出携带者。

其他在珠蛋白编码区发生的突变可分为移码突变，无义突变和通读突变。移码突变与无义突变如果发生在编码区域的3′端就会产生变异的珠蛋白，如果发生在5′端就会由于严重的平截与功能丢失，产生地中海贫血。Cranston血红蛋白包含一变异的β-链，其3′端的移码(在密码子144与145之间两个核苷酸的插入，GA)引起了终止密码子的通读。产生的多肽链比正常的长10个残基。Constant spring血红蛋白由一个通读突变产生，此突变使α-珠蛋白链的终止密码子从UAA变成CAA。这种变异型比正常的α-珠蛋白链长31个残基。

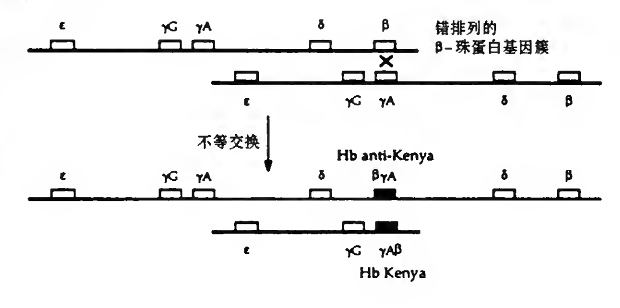
重复DNA序列重组后产生的变异珠蛋白　重复序列之间的错排序列交换(不对称交换，或不对称姐妹染色单体交换)既可能产生在单个珠蛋白基因中的小的重排，也可能产生包括整个珠蛋白基因簇的大的重排。位于β-珠蛋白的第91～94位密码子与第96～98位密码子之间序列GCTG-CACGTG，其同向重复拷贝之间偶尔会发生不对称交换。这导致了一条链的15个核苷酸的缺失与另一条链的同样序列的插入，这样仍然保持了阅读框架，产生了变异珠蛋白Gun Hill。整个基因的错误排列导致的不对称交换产生杂种珠蛋白融合链。最常见的重排包括γ^A-珠蛋白与β-珠蛋白之间的不对称交换，产生血红蛋白Kenya(具有γ^A-珠蛋白的N端与β-珠蛋白的C端区域，见下图)。这个事件还使δ-珠蛋白基因缺失并引起了遗传性的胎儿血红蛋白的延续。不同的Hb Kenya亚型反映了交换的不同的可能位点。对于每一个Hb Kenya染色单体，都有另一个带有互补交换产物的染色单体：Hb反Kenya带有$\beta\gamma^A$-珠蛋白的融合链(此染色单体上还包含一复制的δ-珠蛋白基因)。

引起地中海贫血的突变　地中海贫血源于珠蛋白基因失去表达，此可能由大范围的缺失或更敏感的突变所引起。血红蛋白的可溶性与携氧能力依赖于分子中α-与β-珠蛋白数目的化学计量关系。在α-地中海贫血中，只能提供β-珠蛋白；在β-地中海贫血中，只能提供α-珠蛋白。在这两种情况中，仅存的珠蛋白仍试图形成四体，但这种四体缺少氧载体并且会形成不溶的复合物。

因为在体内有两个多余的α-珠蛋白基因，所以只有当三个甚至全部四个等位基因都丢失的情况下，才会发生严重的α-地中海贫血。大部分α-地中海贫血是由α-珠蛋白基因对之间的不

等交换或染色单体断裂造成的缺失所引起的。偶尔,基因功能的丢失是由一更加敏感的突变造成的,如移码突变,但这在β-地中海贫血中更加常见,因为只有一个β-珠蛋白基因。很多情况下β-地中海贫血是由点突变引起的。包括在β-珠蛋白基因启动区的突变,5′无义突变和移码突变,在多聚腺苷化位点发生的突变与在内含子中发生的突变会阻止剪切。更不寻常的例子包括在β-珠蛋白基因的密码子26位的错义突变,使谷氨酸变成赖氨酸。虽然这是一个非保守的突变,但推测它应产生一变异的血红蛋白而并非引起地中海贫血。然而,这个突变又在β-珠蛋白基因中引入了一个潜在的剪接位点,引起了错误剪切的发生,使野生型的β-珠蛋白的量下降到正常水平的50%~60%。由于点突变或染色体变异影响单个珠蛋白基因产生单地中海贫血,而珠蛋白基因座控制区域(参阅)的缺失会导致多地中海贫血,因为这种区域负责基因簇中所有基因的高水平的协调转录。如β-珠蛋白LCR的缺失引起εγδβ-地中海贫血。

遗传性的胎儿血红蛋白的持续出现(HPFH)在正常成体中,胎儿血红蛋白占血红蛋白总量的1%以下,因为在发育过程中存在γ-珠蛋白基因表达向β-珠蛋白的转变。然而在HPFH病例中,正常发育的转换失效导致HbF占细胞中血红蛋白总量的20%~100%。HPFH的缺失形式与β-地中海贫血同时发生,即部分或全部β-和δ-珠蛋白基因及其控制元件的缺失使正常的转换无法发生,在这种情况下,HbF可以补偿成体血红蛋白的缺少从而导致HPFH(然而,并非所有的缺失型的β-地中海贫血中都发生HPFH)。另一些HPFH的例子与缺失无关,反映了启动子中使转换不能正常发生的突变。另外,也有报道HPFH的病例,未发现在β-珠蛋白的基因簇中有突变,这证明存在基因簇外的调节元件。



文框 15.2 人类疾病中的致病的三密码子重复

三密码子重复综合征 很多人类疾病,包括亨廷顿氏舞蹈症,肌强直营养障碍及脆性X染色体综合症都表现出复杂的家谱,其特征如:不完全外显,变化的表现度及前发性(anticipation)(遗传疾病在延续的子代中有症状严重并且发病时期提早的趋势)。在1990年,对脆性X染色体综合症基因FMR1的序列分析辨明了5′非翻译区的一串联重复三密码子区域,(CGG)*n*的多态区域。在正常的个体中,重复拷贝数目少于50,而在患者中,此区域延伸并且包含超过1000的拷贝。从那时起,10种另外的三密码子重复综合征(tripet repeat syndroms)被鉴定。

三类三密码子重复综合征 脆性X染色体综合征是典型的II类三密码子重复综合征,其三密码子重复出现在非编码DNA区域并且富含GC。致病的三联重复的延伸数目的变化较大(从正常的50~100拷贝到超过2500拷贝),这是通过染色质的再塑及DNA甲基化抑制了转录产生的。亨廷顿氏舞蹈症是I类三密码子重复综合征:重复的三密码子(CAG)*n*位于基因的编码区,编码一串谷

氨酰氨残基。致病的三联重复的延伸数目的变化较小(从正常的 10～30 拷贝到 40～100 拷贝)。转录不受延伸的影响,突变的效应可能作用于蛋白功能的水平上。在肌强直营养障碍症中,在 3′UTR 中发现有(CTG)*n* 重复,数目从小于 50 拷贝到超过 3000 拷贝。它似乎既不影响转录也不影响多肽链结构,其致病机制可能是通过中和了能在转录中能识别三密码子重复顺序的特异 RNA 结合蛋白。已鉴定到几种 CUG 结合蛋白,延伸的重复序列可能与它们结合并阻止它们与正常情况下的靶 RNA 序列结合。

重复延伸的机制　基因内串联重复的不稳定本质正好解释了在三密码子重复综合征中观察到的复杂的家谱情况。有可能重复数目较小的变化是由 DNA 复制中的链滑动引起的,但较大的重复数目增加可能涉及不同的机制,如不等交换或姐妹染色单体不等交换。前发性反映了一旦重复数目达到一严格的**前突变**(premutation)阈值就会发生重复延伸的强烈倾向。在肌强直营养障碍症中,这个阈值处于 40～50 拷贝的较窄的范围中,而在脆性 X 染色体综合征,阈值是 50～200 拷贝。一旦重复延伸到此拷贝数目,就有很大的可能性继续导致致病性延伸,缩短的可能性较低。这种倾向性的本质还属未知。

文框 15.3　随机,定向与编程突变

细菌中随机和定向突变　突变是否仅仅随机发生,为自然选择提供原材料,还是它们是对选择压力作出的反应。突变的随机、非定向的本质于 1943 年 Lucia 和 Delbruck 的**彷徨变异测试**(fluctuation test)首次表明。细菌在优化的条件下培养几代后移入严峻的培养环境(用噬菌体 T1 来感染细菌)。抗噬菌体的突变株在培养几代后分离得到。如果突变是随机发生的(即不依赖于噬菌体感染),一些培养的细菌可能包含较少的噬菌体抗性细胞(因为在实验中突变发生较晚),而另一些可能包含较多(因为突变发生较早,在暴露于噬菌体之前已发生)。相反,如果突变作为对噬菌体感染的反应而发生,所有的培养物应该包含相同数目的抗性细胞。实验结果表明在不同培养物之间抗性细胞的数目有较大的差异,说明突变是随机发生的。

在 1988 年,Cairn 提供证据表明 *lac* - 菌株(乳糖利用的营养缺陷型细菌)在提供乳糖的情况下转变为野生型的速度大大超过正常情况。其他的研究者也在另外的营养缺陷型中发现类似结果。这些结果表明在突变更适合于环境时,细菌可使特定基因发生定向的突变,即突变作为选择压力的反应而发生。自从这些实验的结果发表后,对定向突变的概念有争议,对于这样的过程的机理与进化后果都有热烈的论战发生。

在哺乳类细胞中随机和定向基因扩增　当培养的哺乳类细胞用药物处理来抑制特定的酶时,抗性细胞可以被分离到并生长。抗性的克隆可暴露在渐次增加浓度的药物中,药物的量远远超过起始的细胞群落的致死量,仍然能分离到抗性细胞。例如,药物氨甲蝶呤是叶酸的类似物,它是二氢叶酸还原酶(DHFR)的竞争性抑制剂。野生型细胞对 0.1μg ml^{-1}的氨甲蝶呤敏感,但经过逐步的选择,可以分离到能耐受 1mg ml^{-1}的氨甲蝶呤的细胞,这个量比野生型致死量高五个数量级。对抗性细胞的 DNA 进行分析,发现有些是在 *DHFR* 基因中发生了简单的突变,使酶对抑制剂的敏感性下降,有些突变发生在另外一些基因中,限制氨甲蝶呤的吸收。然而,更多的是扩增了 *DHFR* 基因座,在经过几轮筛选的高抗性细胞中,这个基因座可能扩增上千倍。在这种高选择性的细胞的染色体制备中,可以观察到扩增区域形成一延伸的染色体带(**均匀染色区**,homogeneously staining

region)，或形成小的特别的染色体被称为**双微体**(double minute)。偶尔，扩增的区域会易位到其他的染色体中。

表面上看，当药物剂量上升药物抗性基因扩增的现象似乎是选择压力下定向突变的例子。然而，在第一轮筛选后，在起始的野生型的群落中只有极少数细胞(1:10^7)能存活，这些代表随机发生 *DHFR* 基因座的扩增的细胞。哺乳类基因组的任何区域扩增的自发发生率都很低，而筛选过程富集了培养物中那些已发生有益扩增的细胞。这些细胞一旦被分离并产生抗性克隆，其中的一些细胞可能发生进一步的扩增并经过第二轮药物处理被筛选出来，这个步骤渐次重复进行下去。

扩增单位(**扩增子**，amplicon)往往很大(大于 100kb)并且除了 *DHFR* 基因座以外还包含很多两侧的序列，可能还包括基因组其他部分或其他染色体上的 DNA。重复不是同源的：它们长度不同并且经历了重排。一旦选择压力移去它们变为不稳定。基因扩增可能包括很多混杂和动态的重组事件或期外复制(unscheduled replication)，其机制并未完全了解清楚。非选择性的两侧序列的共扩增可被开发用于哺乳类表达克隆(参见*扩增载体*)。基因扩增作为原癌基因过度表达的主要机制在癌细胞中也可观察到(见第 18 章原癌基因与癌)。

发育中的编程扩增　如同在所有细胞都会发生随机扩增并可以通过药物处理来筛选或在癌中通过体细胞自然选择而筛选到扩增突变体，在发育中特定的扩增事件会以编程方式发生。编程扩增的对象包括很多两栖动物的 rRNA 基因，这些基因以小 DNA 环的形式从基因组中被切离，还有果蝇的卵壳基因会在基因组内选择性地扩增。有关编程扩增的进一步讨论，见第 6 章发育，分子机制。

体细胞超变　体细胞突变的一个清楚的例子是**体细胞超变**(somatic hypermutation)：在 B 细胞发育中，通过向核苷酸序列中导入变化来改变胚系免疫球蛋白 DNA。在人类与小鼠中，体细胞超变特异地发生在免疫球蛋白已经重排与表达的 B 细胞中：这是**亲和力成熟**(affinity maturation)的机理，即抗体与特定抗原的亲和力的提高。在绵羊中，未重排的免疫球蛋白基因发生超变以提供较分散的抗体基本的所有组成成分(repertoire)。由于较低级的脊椎动物具有很少组合与连接的多样性但仍发生体细胞超变(参见 *V(D)J 重组*)，看来体细胞重组的开始的作用是产生一级多样性。

体细胞超变的机制仍属未知，但已鉴别到一超变募集的同一位点，并有几条线索暗示了它与转录的联系。很多工作集中在小鼠 Igκ 位点，转基因 Igκ 被广泛应用在对这个过程的研究中，因为它们可以作为体细胞超变的底物。Igκ 基因的体细胞超变的区域起始于重排 V 片断的前导内含子上游区域中，延伸通过 V 和 J 片断并进入 J－C 内含子中(参见*免疫球蛋白基因*)。然而，突变较大程度地被限制在对应于可变区的 DNA 中，在高变区即抗体实际上与抗原接触的部分簇集。共有核苷酸序列 RGYW 被认为是超变募集位点的一部分，因为很多(但非全部)这种位点是区域超变的热点。对丝氨酸密码子分布的研究表明胚系免疫球蛋白通过进化使超变集中于高变区。丝氨酸由两个不相关的密码子家族编码：AGY(超变同一序列的部分)与 TCN(不是超变序列的部分)。在免疫球蛋白位点中，丝氨酸密码子的使用有偏向性，AGY 密码子倾向于出现在高变区，而 TCN 密码子在别处。而在不发生超变的 T 细胞受体基因中，TCN 在 V 区的各处出现。

超变区域位于 Igκ 转录单位中，且超变显示了独特的链极性，这些数据表明超变与转录相偶联，可能以*转录偶联 DNA 修复*(参阅)相同的方式进行。与转录的联系的进一步的证据来自转基因实验，表明了 κ 轻链增强子是超变所要求的，但启动子与 V 片断的大部分可被异源序列替代仍可作为超变的底物。在超变中具有明确作用的反式因子还未被鉴别，一种可能的候选物是 TFIIH——基本的转录因子，在转录偶联 DNA 修复中起核心作用。现有的模型认为 TFIIH 可吸引易出错的 DNA 聚合酶到超变位点，在下一轮的 DNA 复制中导入核苷酸替代。

文框 15.4 第二位点突变

抑制与增强　**第二位点突变**(second site mutation)是在初始的**第一位点突变**(primary site mutation)以外发生的突变,它可以改变由第一位点突变决定的表型。当第一突变的效应被第二突变所改善,这种现象被称为**抑制**(suppression);当效应被加强,这种现象被称为**增强**(enhancement)。第一位点突变也可被环境所抑制,例如,链霉素可通过减少翻译的忠实度模拟*信息抑制*(见下文);这被称为**表型抑制**(phenotypic suppression)(另见*表型*)。在表型水平上,抑制的效应等同于回复突变。然而只有在抑制中,效应的成分才可被重组即突变位点之间的交换所分离。

在同一基因中的抑制因子　**基因内抑制**(intragenic suppressors)或**内源抑制**(internal suppressors)指第二位点突变发生在第一位点突变的同一基因中(参考*等位互补*),它表现为顺式作用通过使某些结构缺陷改良来保持野生型表型,如在某处发生了单个核苷酸的插入导致移码的第一突变,附近的单个核苷酸的缺失就可作为抑制因子来恢复原来的阅读框。在*密码子内抑制*(intracondon suppressors)的特殊例子中,第二位点突变与第一位点突变发生在同一密码子中,通过恢复密码子原来的含义或将非保守突变改变为保守突变来补偿第一位点突变。

在不同基因中的抑制因子　**基因间抑制**或**基因外抑制**或**外抑制基因**(intergenic or extragenic or external suppressors)指第二位点突变与第一位点突变发生在不同基因中。基因外抑制的效应不是基因水平的补偿,而是产物水平的补偿(即它们反式抑制功能)。在某些例子中基因外抑制因子可能进行生理性补偿(例如,无法合成一种重要的酶如酪氨酸合成所需的酶的功能缺失突变,可能被另一基因中发生的能更有效地从环境中吸收酪氨酸的第二位点突变所补偿)。在另外的情况下,基因外抑制因子可用于鉴别与其相互作用蛋白的基因,筛选非联系的抑制因子突变体已广泛用于这个目的(另见*双杂交系统*)。基因外抑制是*非等位相互作用*的一种形式(参阅)另见互补。

信息抑制因子　**信息抑制因子**(informational suppressors)或称**超抑制因子**(supersuppressors)通过在相应的 tRNA 的*反密码子环*(参阅)内引入突变来补偿错义、无义甚至小的移码突变的基因间抑制因子,这样使突变的编码区域仍然按野生型的密码解读。无义抑制因子根据它们改变为有义密码子的无义密码子的类型被分为**琥珀抑制因子**(amber suppressors),**乳白抑制因子**(opal suppressors)和**赭石抑制因子**(ochre suppressors)。因为 tRNA 基因一般出现在很多拷贝中,一个信息抑制因子的出现不会导致所有终止密码子的错误替换;因此野生型基因的终止正常发生并且机体可以存活。

增强突变　第二位点突变增加了原来突变表型的严重程度,这个过程被描绘为**增强**(enhancement),这种突变被称为**增强突变**(enhancer mutation)。如同抑制因子可能是顺式或基因内[1]的,也可能是反式或基因间的,后者可鉴别相互作用的基因产物。增强突变不可与*增强子*(参阅)一种顺式作用调节元件混淆。

(徐　沁 译)

1)原文误写为"基因间的"(intergenic),应为基因内的(inteagenic)。——译者注

第16章
核 酸 结 构

基本概念及定义

- DNA 及 RNA 被称为**核酸** (nucleic acids),是由核苷酸单体构成的多聚体。核苷酸是含氮碱基连接到磷酸化的糖环上形成。糖环由 5′→3′磷酸二酯键相连接,形成一条固定但有极性的主链,碱基伸出于主链之外。
- 核酸多聚体携带碱基的性质及其排列顺序组成了遗传信息。特定碱基间可相互作用形成互补配对,使核酸具有双螺旋结构、可充当模板及识别同源序列。并以此为基础进行复制、重组和基因表达 (参阅)这些重要的生命过程。
- 双螺旋核酸的构象依赖于碱基序列,拓扑限制,不同的外界环境及与蛋白质相互作用。这种**构象多态性**(conformational polymorphism)与碱基序列本身的多态性一样,对核酸的功能有重要作用。
- DNA 是细胞的遗传物质,它主要以双链形式存在 - 这使它特别适合作为遗传物质的载体,即信息的**蓝本** (blueprint),因为它能够通过模板的自我修复保持遗传信息的完整性 (见第 14 章诱变和 DNA 修复) 。细胞内的 RNA 从 DNA 转录而来,它主要以单链形式存在,通过折叠形成复杂的二、三级结构。胞内存在几类 RNA,分别行使不同的功能,主要参与遗传信息的表达(表 16.1)。病毒的基因组既可以是 DNA 也可以是 RNA (见第 30 章病毒)。

16.1 核酸的基本结构

核苷酸的结构 核苷酸是核酸的基本重复单位,由碱基、糖及磷酸三部分组成。核苷酸在细胞内有多种功能,如:作为能量的载体,神经递质或作为第二信使(见信号传导)。

碱基(bases)是碱性含氮杂环化合物**嘧啶** (pyrimidine)及**嘌呤** (purine)的衍生物(图 16.1)。DNA 和 RNA 都含有四种**主要碱基** (major bases),其中**腺嘌呤** (adenine)、**鸟嘌呤** (guanine)及**胞嘧啶** (cytosine)是两者共有,此外**尿嘧啶** (uracil)为 RNA 特有,**胸腺嘧啶** (thymine)为 DNA 所特有。DNA 含有胸腺嘧啶可能是进化过程中形成的,它避免了由于胞嘧啶脱氨基形成尿嘧啶产生的突变(参见 5 - 甲基胞嘧啶)。在 DNA 及 RNA 中都含有**稀有碱基**(minor bases, 如次黄苷),它可直接来源于聚合过程,也可能是聚合后修饰的结果。(参见 DNA 修饰,tRNA,RNA 编辑,碱基类似物)。碱基可以多种异构体形式存在,它们形成氢键的能力各不相同,因而导致各种突变的产生(见第 14 章诱变和 DNA 修复)。DNA 及 RNA 中的常见碱基通常处于一种相对稳定的异构体状态(主要异构体形

式,dominant tautomeric form),因此它们才被用以携带遗传信息。

表 16.1 细胞内 RNA 主要与次要的功能类型

RNA 的种类	功能
主要类型	
mRNA(信使 RNA)	转录自编码蛋白质基因的 RNA,携带翻译信息。有些类 mRNA 的转录本不能被翻译,如:XIST,H19 (参见亲本印迹)
hnRNA(核内不均一 RNA)	mRNA 剪接前体。真核基因经修饰的转录产物。由于它的大小与 tRNA 及 rRNA 相比变动范围很大,因而得名
tRNA(转运 RNA)	在翻译过程中的转接分子。在反转录病毒复制的过程中,tRNA 可作为 DNA 复制的引物(参见反转录病毒)
rRNA(核糖体 RNA)	是核糖体的主要结构组成部分,蛋白质合成过程所必需。
次要类型	
iRNA(起始 RNA)	在 DNA 合成中作为后滞链合成引物的短 RNA 片段(参见复制)
snRNA (核内小 RNA)又称 **U-RNA**(富含尿嘧啶的 RNA)	核质中的低分子量 RNA,富含经修饰的尿嘧啶残基,参与内含子的剪切及其他加工过程
snoRNA(核仁小 RNA)	见于核仁的低分子量 RNA,可能参与了 rRNA 的加工过程
scRNA(胞质内小 RNA)	在胞浆中发现的有多种功能的低分子量 RNA。例如:7SRNA 是信号识别颗粒的组成部分(参阅),而 **pRNA** (prosomal RNA) 是能与近 20 种蛋白质相互作用的 RNA 分子,并且发现在 mRNP 或称信息体(参阅)中,它能与 mRNA 共同被包装,可能对基因表达有普遍的调控
端粒酶 RNA	作为形成端粒 (参阅)重复序列的模板的核 RNA,是端粒酶(参阅)的组成部分
gRNA(指导 RNA)	在锥虫动基粒中合成的 RNA 种类,可作为 RNA 编辑过程中的模板
反义 RNA (mRNA 干扰互补 RNA, micRNA)	反义 RNA 与 mRNA 互补,可与其形成双螺旋结构阻断蛋白质的合成。自然产生的反义 RNA 见于多种生物系统,但在细菌中最常见。可被称为 mRNA 干扰互补 RNA(参见质粒复制,F 转移区域,λ 噬菌体,蛋白质合成调控及基因治疗)
核酶	指具有化学催化功能的 RNA 分子(RNA 酶)。它通常具有自我催化功能(参见自身剪切内含子),而核糖核酸酶 P 是真正意义上的催化剂 (参见 tRNA 的加工)。其他的 RNA 与蛋白质共同发挥作用,例如:在线粒体 DNA 复制中的 MRP 内切酶(见第 19 章细胞器基因组)

大多数 RNA 是线性的,但也有分支状的(如:在内含子加工过程中的套索结构),也有环状的(如:类病毒及 SRY mRNA;参见性别决定)。

DNA 和 RNA 都含有五碳糖(sugar，**戊糖**,pentose),其糖分子内部产生的半缩酮基团使之形成了**呋喃糖**（furanose)环结构［由于这个结构与杂环化合物**呋喃**（furan)相似］(图 16.1)。DNA 与 RNA 本质的不同在于它们含有的五碳糖类别不同:RNA 含有 D-核糖(D-ribose)［因而称为**核糖核酸**(ribonucleic acid，RNA)］,而 DNA 含有 **2′-脱氧-D-核糖**（2′-deoxy-D-ribose),即核糖 2′位的羟基被氢原子取代(因而称为**脱氧核糖核酸**,deoxyribonucleic acid ,DNA)。这一结构上的微小差异使 DNA 和 RNA 的理化性质有很大不同。DNA 由于立体化学的空间位阻（steric hindrance)而更加坚硬,而且在碱性条件下不易水解。这部分解释了为什么 DNA 被选作基本的遗传物质。

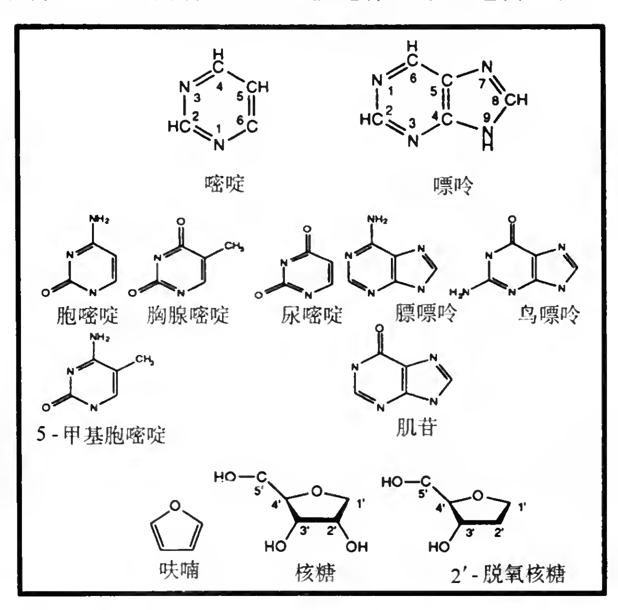


图 16.1 核酸中的碱基及糖。核酸中常见的几种碱基,胞嘧啶、胸腺嘧啶及尿嘧啶是杂环化合物嘧啶的衍生物,而腺嘌呤及鸟嘌呤是杂环化合物嘌呤的衍生物。胸腺嘧啶和尿嘧啶结构相似,它们都能以相似的方式与腺嘌呤配对(参见*碱基互补配对*)。所以 DNA 中的胸腺嘧啶在 RNA 中可为尿嘧啶所替代。图中也显示了 DNA 中的稀有碱基:5-甲基胞嘧啶及肌苷。D-核糖及 2′-脱氧-D-核糖被称为呋喃糖,因为在结构上它们与杂环化合物呋喃相似。图中标出了常规的环上原子的记数法,为了与碱基的记数系统相区别,糖环上的记数加了“′”。

碱基连接于五碳糖的 $C^{1'}$ 位形成**核苷**(nucleosides)。糖的 $C^{1'}$ 碳原子与嘧啶的 N^1 氮原子相连接或与嘌呤的 N^9 氮原子相连(图 16.2),形成 **β-N-糖苷键**（β-N-glycosidic bond)。核苷的命名法与碱基有一些不同(表 16.2)。

核苷酸(nucleotides)是核苷的磷酸酯。酯化可发生在核苷酸的任何游离羟基,但在核酸中最常发生在 5′和 3′位点。磷酸基通过**磷酸一酯键**（phosphomonoester bond)与糖环相连,再通过**磷酸苷键**（phosphoanhydride bonds)可顺序连接多个磷酸基团(图 16.2)。5′-三磷酸核苷是核苷酸合成时的底物。两个羟基可通过同一磷酸基团酯化形成**环化核苷酸**（cyclic nucleotide),例如:环腺苷酸（cAMP，3′-5′-环磷酸腺苷;见第 28 章信号转导)。

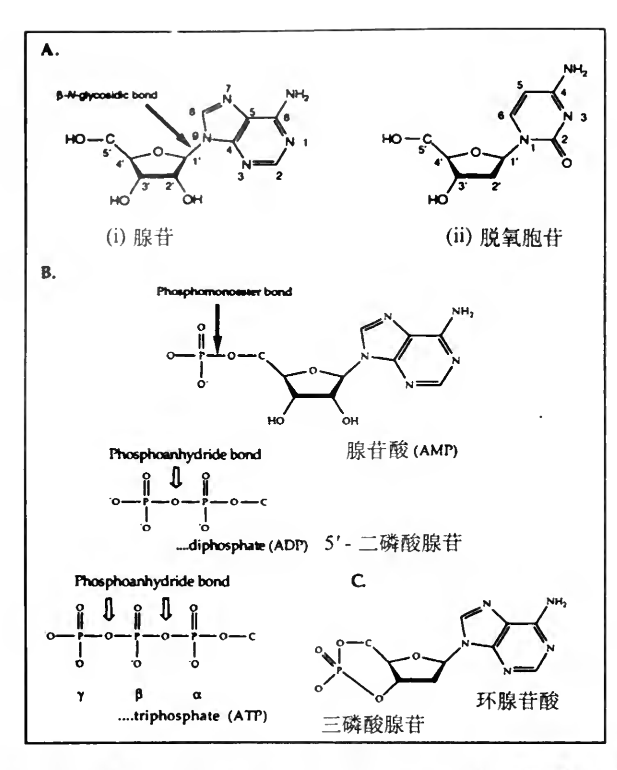


图 16.2 核苷和核苷酸。(a) 两个核苷的化学结构图中显示了连接糖与碱基的糖苷键。(i) 腺嘌呤核苷:含嘌呤的核糖核苷。(ii) 脱氧胞嘧啶核苷:含嘧啶的脱氧核苷。(b) 单磷酸腺嘌呤核苷酸(腺苷酸)及其二磷酸,三磷酸衍生物,图中显示磷酸基团的命名法。并标明了单酯键及磷酸苷键的位置。(c)环状核苷酸:环腺嘌呤单磷酸。

表 16.2 碱基及其核苷、核苷酸衍生物的命名规则(括号内为缩写)

	DNA	RNA
碱基	腺嘌呤(A)	腺嘌呤(A)
	胞嘧啶(C)	胞嘧啶(C)
	鸟嘌呤(G)	鸟嘌呤(G)
	胸腺嘧啶(T)	尿嘧啶(U)
核苷	脱氧腺苷(dA)	腺苷(A)
	脱氧胞苷(dC)	胞苷(C)
	脱氧鸟苷(dG)	鸟苷(G)
	(脱氧)胸苷(dT)	尿苷(U)
核苷酸(以腺嘌呤衍生物为例)		
一磷酸核苷	5′-单磷酸脱氧腺苷 (dAMP, dpA)	5′-单磷酸腺苷或称腺苷酸(AMP, pA)
二磷酸核苷	5′-二磷酸脱氧腺苷 (dADP, dppA)	5′-二磷酸腺苷 (ADP, ppA)
三磷酸核苷	5′-三磷酸脱氧腺苷 (dATP, dpppA)	5′-三磷酸腺苷 (ATP, pppA)
核苷酸残基	脱氧腺苷酸	腺苷酸

在缩写中,脱氧核糖的核酸及核苷酸衍生物的区别在于前缀'd'。当需要特别说明时,核苷及核苷酸前缀为'r',例如 ATP=dATP。只有第二个缩写字母可区分5′及3′磷酸,5′磷酸残基置于碱基之前(如 pA 是指5′单磷酸腺苷),3′磷酸残基置于碱基之后(如 Ap 是指3′单磷酸腺苷)。胸腺嘧啶衍生物前的脱氧前缀可以省略,因为它是一种 DNA 特异性的碱基,它所含的糖显然是脱氧核糖。由于胸腺嘧啶很少出现于 RNA 中,这种表示方法也是一种习惯用法。当上下文很清楚的情况下,DNA 及 RNA 的序列都可以用碱基序列来表示。

不确定碱基表示法:R=A 或 G(任何嘌呤); Y=C 或 T/U(任何嘧啶); K=G 或 T; M=A 或 C; S=G 或 C(三键,强相互作用); W=A 或 T(二键,弱相互作用); B=G,T 或 C(即非 A); D=G,A 或 T(即非 C); H=A,C 或 T(即非 G); V=A,C 或 G(即非 T)。

核酸的一级结构　　**核酸**(nucleic acids)是核苷酸单元组成的长链,或称**多聚核苷酸**(polynucleotides),多聚反应的底物是三磷酸核苷,但核酸中的重复单位,或称为单体,是单磷酸酯化的(一磷酸核苷残基,核苷酸化残基,核苷酸残基)。在聚合过程中,链末端核苷酸残基的3′羟基对即将掺入核苷酸的α-磷酸(也就是最内部的磷酸)进行亲核攻击形成一个5′→3′**磷酸二酯键**(5′→3′phosphodiester bond)。这一反应由DNA或RNA聚合酶催化(box 26.1),同时形成副产物焦磷酸(参见DNA复制,转录)。多次聚合反应产生链状聚合物,它包括固定的具有**5′→3′极性**(5′→3′ polarity)的**糖-磷酸主链**(sugar-phosphate backbone)以及向外伸出的含氮碱基。DNA和RNA的基本化学结构及常用简写形式见图16.3(另参见RNA)。

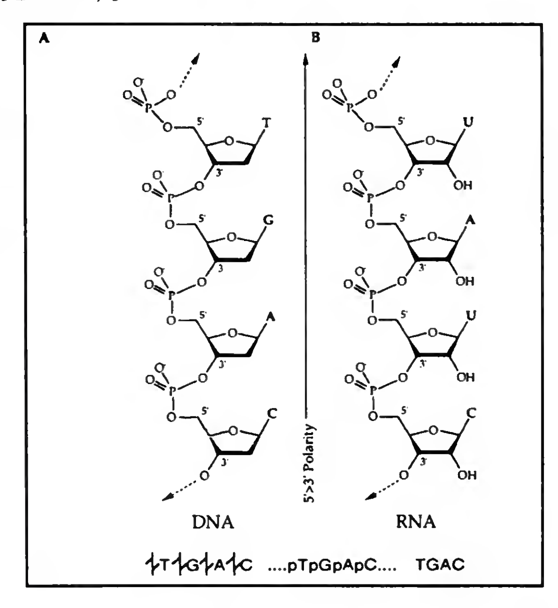


图16.3　一小段DNA(a)及RNA(b)及基本化学结构,图中标示出了5′→3′磷酸二酯键,并显示了三种一级序列缩写方法。

寡核苷酸(oligonucleotides)指较短的核酸(如长度小于<100nt)。寡核糖核苷酸在细胞中天然存在,它可作为细胞内DNA复制的引物,也可有其他用途。可用化学方法合成寡核苷酸,它在实验上有很多用途(例如并参见DNA测序,聚合酶链式反应,原位杂交,核酸探针,核酸杂交及基因治疗)。

16.2　核酸的二级结构

碱基间相互作用的两种方式　　通过碱基间两种非共价结合力:碱基配对及碱基堆积,形成了**核酸的二级结构**(nucleic acid secondary structure)。**碱基配对**(base pairing)是一种氢键结合力,也是使核酸链相互缔合的主要作用力,另外通过相邻碱基π电子形成的疏水作用力使整个结构保持稳定,被称为**碱基堆积**(base stacking)作用力。DNA的二级结构是指通过分子间相互作用形成的**双链**(double-stranded)或称**双螺旋**(duplex)分子

(dsDNA)。RNA 分子通常以单链形式存在,它的二级结构反映了分子内碱基间的相互作用。

碱基互补配对　Edwin Chargaff 最早指出,尽管从不同生物体提取的 DNA 中四种碱基的丰度有很大差别,但腺嘌呤与胸腺嘧啶的比值和鸟嘌呤与胞嘧啶的比值总是相同的。**Chargaff 定律**(chargraff's rules) 说明 A = T 和 G = C。Chargaff 定律的分子机制是双链 DNA 中 A 与 T 以及 C 与 G 间的**碱基互补配对** (complementory base pairing)。碱基间通过**氢键** (hydrogen bonds)形成特殊、稳定的构型从而促进配对的形成。一条链上的嘌呤总是与另一条链上的嘧啶配对,这样就形成了由 Jame Watson 和 Francis Crick 提出的 DNA 双螺旋结构 (见下文)。氢键的供体及受体的相对位置导致了特异配对的形成。A:T碱基对间存在两个氢键(供体 - 受体 对受体 - 供体),C:G 碱基对间为三个氢键(受体 - 受体 - 供体 对供体 - 供体 - 受体)——见图 16.4。**Watson-Crick 碱基配对**(Watson-Crick base pair)是核酸二级结构形成的基础,并解释了 Chargaff 定律,同时它也说明了为什么 DNA 能够作为*转录*及*复制*的模板(参阅),能够进行*重组*(参阅)以及通过*修复*(参阅)维持遗传信息的完整性。在 RNA 中,尿嘧啶替代了胸腺嘧啶,由于尿嘧啶与胸腺嘧啶有着相似的化学结构,它能与腺嘌呤形成同样的氢键配对。两种核酸杂交遵循同样的规律。虽然这类相互作用十分广泛,但还存在其他的碱基配对方式,在形成核酸二、三级结构时发挥着重要作用。这将在下文中讨论。

图 16.4　Watson-Crick 碱基配对。氢键位置用虚线表示。G - C 对间形成三个氢键,A - T(或 A - U)对间形成二个氢键。因而 G - C 对更加稳定(参见 GC - *含量*,*热解链曲线*)。

其他形式的碱基配对　Watson-Crick 碱基配对对于核酸结构维持及功能的发挥至关重要。然而,碱基间两个氢键有 28 种可能的排列方式,这使得碱基间的相互作用也可有多种形式。在各种可变构型中,最有意义的是 **Hoogsteen 碱基配对** (Hoogsteen base pairs),它在 tRNA 结构及三螺旋结构的形成中起重要作用。**摆动配对** (wobble pairs)是对 Watson-Crick 碱基配对的重要补充,它允许 tRNA 反密码子上 5′端第一个碱基与 mRNA 中的碱基不完全配对(参见*遗传密码*,*摆动假说*)。碱基在 Watson-Crick 碱基配对中的位置发生偏移,失去一个氢键后产生了摆动配对现象。

DNA 双螺旋　1953 年,Jame Watson 和 Francis Crick 通过 X 射线衍射实验研究(由 Maurice Wilkins 及 Rosalind Franklin 提供 DNA 纤维标本),解析了 **DNA 双链**(double-stranded DNA)结构 (dsDNA)。双链 DNA 通常由两条**反方向平行** (antiparallel)的单链(即 5′→3′的极性相反)围绕一公共轴旋转形成右手双螺旋 。由糖 - 磷酸重复单位构成了螺旋的**主链**(back bone),碱基向螺旋的内部伸展、对应的碱基间可形成穿过螺旋中轴

的氢键。由碱基配对及相邻碱基间(π－π)堆积力维持双螺旋结构的稳定。碱基位于螺旋内部,受外界环境的影响较小(使遗传信息免受各种理化因素的影响,见突变及DNA修复)。由于碱基间的相互作用使得螺旋的表面并不是光滑的,而是有较宽的**大沟**(major groove)及较窄的**小沟**(minor groove)相间隔,这有利于序列特异的蛋白质与DNA间的相互作用(见第17章核酸结合蛋白)。

表 16.3 三种主要DNA螺旋类型的形态特征、扭转角及螺旋参数的比较

	构象		
	A	B	Z
形态特征			
螺旋方向	右	右	左
螺距(每圈碱基数)	11	10	12
大沟	深,窄	宽	平
小沟	宽,浅	窄	窄,很深
螺旋直径	2.3nm	1.9nm	1.8nm
扭转参数			
糖环构象	$C^{2'}$内式	$C^{3'}$内式	交替
糖苷键角	反式	反式	反式和顺式交替
螺旋参数			
位移	－4.4	0.6	3.2
扭转	33	36	－49/－10
上升	2.6	3.4	3.7
倾斜	22	－2	－7

由于B－DNA中的位移及倾斜很小,其中的碱基与螺旋轴是垂直的(图16.5)。

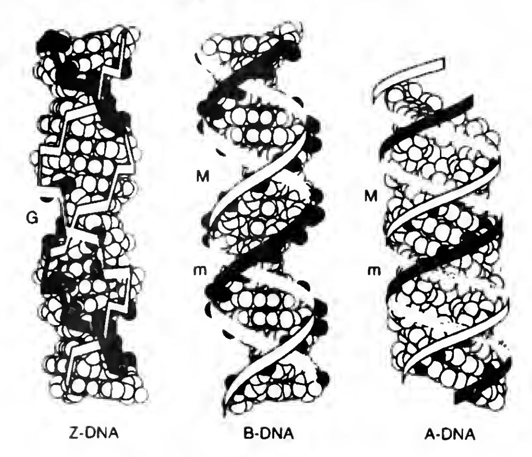


图 16.5 A-DNA,B-DNA及C-DNA双螺旋的结构。在体内,B-DNA形式占绝大多数。M指大沟,m指小沟,G指Z－DNA中的单一的沟。(引自:DNA复制,Kornberg及Bader(1992),WH Freeman,New York。)

以上DNA双螺旋的结构特征是由长的DNA纤维得到的,因而是普遍的性质,没有考虑由特殊碱基序列结构引起的局部的扰动。而且除了被称为B-DNA的最常见的双螺旋结构,体内也存在着不同结构的螺旋形式(见下文)。

多种螺旋构象　对DNA二级结构的研究表明,在不同的湿度下DNA可呈现不同螺旋

构型(**异构体**,conformers)(通常用大体形态,键角及螺旋参数来描述螺旋构象,见文框16.1)。**A-DNA** 及 **B-DNA** 都为右手螺旋,但 A-DNA 较宽,每圈螺旋含有更多碱基,而且比上面描绘的经典的 B-DNA 刚性强。这反映了螺旋的盘绕方式及碱基平面倾角的差异,以及由于脱水作用使碱基对离开了螺旋轴。构型的改变使螺旋表面大沟、小沟的形状也起了变化,从而影响了蛋白质与 DNA 间的相互作用。在生理状况下,双链 RNA 和 RNA:DNA 杂交体都处于 A 型结构,因此它们比 DNA 缺乏柔韧性。A 型的 DNA 比 B 型难以溶解,这就是为什么过度干燥的 DNA 会很难溶解,如在提取质粒时会遇到的那样。

由于分析方法的改进以及合成特定寡核苷酸能力的提高(见上文),已经观察到许多不同构型的 DNA 螺旋结构。它们基本上可分为三类,前述的 A 型及 B 型以及 Z 型(表16.3 及图 16.5)。早期对由嘌呤-嘧啶交替组成的寡核苷酸的研究发现了左手螺旋构型的 **Z-DNA**。它的结构特征为以两个碱基对为周期交替出现的螺旋参数与扭转角,这使螺旋的主链成为锯齿状(这也是 Z-DNA 名称的来源)。由此可见,Z-DNA 的重复单位为两个碱基对而不是像 A-DNA 和 B-DNA 中的一个碱基对。尽管由嘌呤-嘧啶交替组成的寡核苷酸,如 oligo-dGdC,oligo-dAdC,都是形成 Z-DNA 的良好底物,但现在还不知道这一序列特异性是否为形成 Z-DNA 的充分条件或必要条件。在过度扭曲的 DNA 中易于形成 Z-DNA 的构型,并由脱水作用而稳定,这在基因表达的调控中可能起重要作用(见转录)。

DNA 结构中的局部可变性 对寡核苷酸晶体的分析表明,与 DNA 长纤维有很大不同,不同的碱基顺序产生差异很大的螺旋参数,这是由于不同的序列影响螺旋与扭转参数以得到最稳定的碱基堆积及配对方式。从这个角度来说,B-DNA 柔韧性较大,由于特异碱基序列可形成各种不同的局部构型。这也提示,DNA 的构型并不是一成不变的,而是可以在不同构型间转换。就某一种特定的构型来说,也存在着局部多态性,如:由于局部多态性而形成的 DNA 弯曲及**螺旋转换**(helical transitions,指在一个分子中出现的由一种螺旋形式突然转换为另一种形式,如:B-Z 转换)。**DNA 弯曲**(DNA bending)是由于局部碱基序列的不同由碱基堆积作用引起的固有的性质,可表现为**各向同性**(isotropic,无偏性)或**各向异性**(anisotropic,因为 DNA 向某特定方向折叠)。在富含 AT 区及与螺旋周期相一致的 GGCC 重复区,DNA 倾向于形成弯曲构型。蛋白质(见核酸结合蛋白)及环化(参见 DNA 拓扑学)都可诱导 DNA 弯曲的形成。在 DNA 组装形成染色体(见染色质)及复制、重组、转录(参阅)的过程中被诱导 DNA 弯曲都是必不可少的。蛋白质可能可识别处于某种特定弯曲构型的 DNA(如拓扑异构酶)。

RNA 及非双链 DNA 的二级结构 在 RNA 及 DNA 的单链区,其二级结构是由分子内的碱基配对决定的。但是细胞内的 DNA 通常以双链形式存在,能够形成分子内配对的碱基很少见。与此相反,在胞内 RNA 中,分子内形成二级结构的情况很常见,并且是它们行使功能的分子基础。在基因的表达及调控过程中,RNA 的二级结构起了主要作用:rRNA 与 mRNA 间的碱基配对控制着蛋白质合成的起始,tRNA 与 mRNA 间的碱基配对促进了翻译过程,RNA 发夹结构及茎环结构控制了转录的终止,翻译的效率以及 mRNA 的稳定性。RNA-RNA 间的碱基配对在内含子的剪切过程中也起着重要作用(见转录,

RNA加工,蛋白质合成各章)。表16.4列出了几种主要的分子内核酸二级结构。与DNA相似,局部碱基序列特征对RNA的构型也有影响,而且,相对较高的修饰碱基含量使RNA的结构更加多变。

表16.4　核酸分子链间二级结构元件

二级结构	定义
凸起及凸环	在双螺旋的一条链上有额外碱基时出现的畸形结构。**凸起**(bulges)是由于多出了单个碱基。凸环是由于多出一个以上的碱基。它破坏了邻近碱基的堆积力,诱导形成弯曲,增加了DNA大沟的柔顺性。双链DNA中的凸起和凸环是由于插入形成的(见第14章诱变和DNA修复)。
内部环(泡)	在双链结构中由于一个或多个碱基错配形成的畸形结构。在RNA中,内部环或泡与蛋白的识别位点有关。
发夹结构,茎环结构	具有不完善的两侧对称结构的区域形成的二级结构。两段互补的序列可配对形成茎干,它可能完全互补形成**发夹**(hairpin),也可能末端有未配对的碱基形成**茎环结构**(stem loop),**发夹-环结构**(hairpin-loop)。由于环内特殊的碱基配对及碱基堆积力,三到四个碱基形成的环特别稳定。在许多生物系统中,发夹结构及茎环结构都是有重要生物功能的元件,包括tRNA和rRNA的结构,转录终止,调控翻译及病毒基因组的包装,说明这一结构是RNA-蛋白的相互作用位点。
柄环结构	一段线型的核酸:其末端是互补的可形成一段短的双链区,分子的其余部分形成环状结构。
十字形结构	在具有两侧对称序列的双链核酸中当两条链都可自发形成发夹或茎环结构时可产生这种十字形结构。这种结构是复制错误的主要来源,单链结合蛋白可抑制这种结构的形成。

套索状结构(lariats参阅)通常被归为二级结构,但由于它们是由核苷酸间的共价键连接形成的,所以严格地说应属于一级结构。

16.3　核酸的三级结构

DNA中的三链相互作用　　**核酸的三级结构**(nucleic acid tertiary structure)反映了对整体三维形状有影响的相互作用,包括不同二级结构元件间的相互作用、单链与二级结构元件间的相互作用以及核酸的拓扑特征。

在DNA中单链与双链及双链与双链之间的三级相互作用形成了三链或四链结构。鸟嘌呤可形成**碱基四聚体**(base tetrads),从而使含有一段多聚鸟苷酸的DNA可形成四螺旋结构,它可能与*端粒*结构的形成有关(参阅)。单链DNA可与双链DNA分子中的碱基发生相互作用,通过双链DNA的大沟第三条链与双链中的一条链的碱基间形成非Watson-Crick配对,从而自发形成DNA三链结构。对寡核苷酸与双链DNA相互作用的研究使人们首次认识到三链螺旋的存在,并且发现了四种常见的包括Hoogsteen碱基配对**碱基三聚体**(base triples)。含有同聚嘌呤/同聚嘧啶序列的DNA可通过Hoogsteen碱基配对形成分子内三链结构,称为**H-DNA**(由于质子化而得名)。形成H-DNA的生理意义目前还不清楚,但有证据表明它与某些基因的调控有关,如哺乳动物中的*GAP*-43(参见*三链螺旋治疗*)。当发生DNA重组时,单链DNA侵入DNA双螺旋也可形成三链

DNA 结构。由于拓扑结构的限制(见下一章),侵入链必须与双螺旋中的一条链配对而不能缠绕于其上,形成一种称为**平行汇接**(paranemic joint)的结构,这种结构必须通过与蛋白质相互作用才能达到稳定状态。这一过程包括双链 DNA 的较大程度的解螺旋或形成左手或右手螺旋交替的片段(**V-DNA**)。如果侵入链有一自由末端或有较强拓扑异构酶活性,则它能与双链中对应的互补链形成正常双螺旋以形成**相缠汇接**(plectonemic joint)结构,而双链中的另一条链则游离出来形成**替代环**(displacement loop,D-loop)。当从双链 DNA 转录 RNA 这一过程被原位固定下来时,可形成类似的被称为 **R-环**的三级结构。这种情况可在 ColE1质粒复制的引发阶段发生(见第 20 章质粒)。在 DNA 重组过程中也可以形成**霍利迪连接体**(Holliday junctions),这是一种由四条链形成的三级结构(参见同源重组)。

RNA 的三级结构 RNA 的单链、环状结构及双螺旋间的相互作用产生了 RNA 复杂的高级结构。例如,在 tRNA 中就有碱基三聚体、部分三链螺旋、**茎连接体**(stem junctions,由两个或多个双螺旋区结合而成)及**假结**(pseudoknots, 由 RNA 单链与茎环结构相互作用而形成)等三级结构。

RNA 的折叠与蛋白质折叠相似,通常是在分子伴侣(参阅)的控制下完成。RNA 只有具有复杂的三级结构才能成为有生物活性的分子。象蛋白质一样,RNA 也能催化特定的化学反应。这种有催化活性的 RNA 称为**核酶**(ribozyme)。一些核酶具有自身催化功能(如:自剪接内含子的转录,参阅)。其余的需要反式激活,包括核糖核酸酶 P 及来自于植物类病毒(参阅)的锤头核酶(hammerhead ribozyme)家族。锤头核酶的名称来源于其催化结构域的三级螺旋结构(另参见基因治疗)。

DNA 的拓扑结构 拓扑学是数学的一个分支,它主要研究物质几何结构的特性,这种特性与形状、大小无关,并且不会由于变形而改变。如果双链 DNA 分子有自由末端,这两条链可按能量最低的原则相互缠绕,分子处于**松弛**(ralexed)状态。在松弛状态时,一条链缠绕另一条链的环数称为**双链盘绕数**(duplex winding number)。如果分子中导入额外的缠绕使双螺旋**过度缠绕**(overwound),此时被称为**连接数**(linking number)的总螺旋环数大于双链盘绕数。相反,如果由于外界环境的作用使双螺旋**缠绕不足**(underwound),此时的双链盘绕数就大于连接数。在这两种情况下,双螺旋都可通过一条链相对于另一条链的旋转来恢复到松弛态。对于闭环双螺旋来说,由于不存在自由末端,连接数就成为其拓扑学特征——它不随形状变化而变化,只有当双螺旋 DNA 断裂成为开环状态时才能改变这一特征。当闭环 DNA 处于过度缠绕或缠绕不足状态时,它只能通过螺旋轴自身的盘绕形成**超螺旋**(supercoiling)来缓解扭曲的张力。超螺旋是核酸的四级结构的另一种形式,它主要涉及扭转张力对双螺旋形状的影响而不是链之间的相互作用(文框 16.2)。

超螺旋的生理意义在于自由状态的 DNA 通常是没有生物活性的。许多重要的生物过程需要引入负超螺旋,如:复制,转录及重组的过程。超螺状态的 DNA 储存了驱动这些反应所需的能量。真核细胞的染色体是线性的,通过支架蛋白将两端固定于结构蛋白,然后染色质自身盘绕导入拓扑学张力。核小体的结构使真核细胞的 DNA 维持负超螺旋形式(参见染色质环,基质结合区域,DNA 拓扑异构酶,位点专一性重组)。

核酸的四级结构　在许多结构中核酸以反式作用(如:核糖体,剪接体),这可看作是核酸的四级水平的结构。核酸还可与多种蛋白质相互作用(如:基因组结构蛋白,转录因子,酶及剪接因子)。这些蛋白质的大部分对 DNA 或 RNA 的构象有显著影响。与蛋白质的相互作用可能是序列特异的,也可能是普遍的,并可能引起 DNA 结构或多或少的改变。例如,限制性内切酶 *EcoR* Ⅰ及 *EcoR* Ⅴ都可在其识别序列处引起 DNA 明显的纽结,这可能对其内切酶活性有促进作用。HMG 家族的蛋白质可使与其结合的 DNA 弯曲从而有利于结合于 DNA 不同位置的蛋白质间的相互作用。关于核酸与蛋白质间的相互作用的进一步讨论见第 17 章核酸结合蛋白。

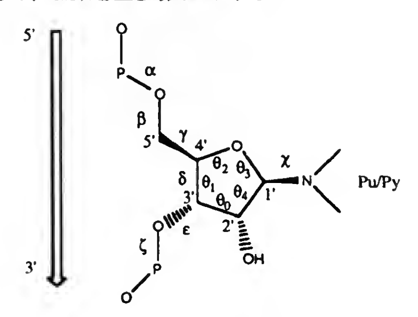
文框 16.1　螺旋的形态,参数及扭转角度

螺旋的形态　DNA 和 RNA 螺旋根据其大致形态来分类。其标准包括螺旋直径,**螺旋方向**(helical sense),**螺距**(pitch, 指每一圈的碱基数),及大,小沟的宽度和深度。

大沟和小沟起初是根据它们在 B-DNA 中的相对大小来定义的。然而,由于在构象紧张的情况下,它们的大小是可以改变的,所以还需要有更精确的定义。大沟被定义为含有嘧啶的 C^4 或嘌呤的 N^7,而小沟包含嘧啶的 O^1 或嘌呤的 N^3。

扭转角度　构象结构的细节可用糖－磷酸骨架,呋喃环自身及糖苷键的各种扭转角度来清楚地描述,可用以下五个键对构象进行简单描述:$\delta(C^{3'}-C^{4'}$,与糖基平面有关的键),χ(糖苷键),$\gamma(C^{4'}-C^{5'})$及磷酸键 α 和 ζ。另外,这些键大部分不是彼此独立的,因而可以不用绝对的键角来表示,而只说明其大概的构象,例如用"反式"或"顺式"来表示糖苷键,用 C^3 内或 C^2 内来表示糖环的位置。

螺旋参数　可用一套通用的**螺旋参数**(helix parameters)来描述堆积或成对碱基间的位置关系。用相对位移及旋转的参数可描述在配对碱基中的碱基相对位置,在堆积碱基中相邻碱基间相对位置及碱基对相对螺旋轴 的绝对位置,这些参数见下表。



坐标	X 轴	Y 轴	Z 轴
相对螺旋轴的绝对位置	X 位移	Y 位移	
逐个碱基对的取代	移位(shift)	滑动(slide)	上升(rise)
取代碱基对中的碱基	切变(Shear)	拉长(stretch)	交错(stagger)
相对螺旋轴的绝对扭转	倾斜角(Inclination, η)	倾斜(Tip, θ)	
逐个碱基对的扭转	倾斜(Tilt, τ)	摇摆(p)	扭转(Ω)
配对碱基间扭转	弯曲(κ)	螺旋扭转(ω)	展开(σ)

文框 16.2 DNA 的拓扑性质的定量表示

螺旋缠绕数的量度　　**连环数**(linking number, L)指双螺旋中两条链相互缠绕的次数,在右手螺旋中 L 为正数。**双链缠绕数**(duplex winding number, L_0) 指松弛态 DNA 的连环数,它代表了 DNA 最低能量状态的构象。对 B－DNA 来说,平均 $L_0 = n/10.3$,其中 n 为碱基对。如果 DNA 为松弛态则 $L = L_0$。任何使 DNA 过度缠绕或缠绕不足的情况都会产生扭曲张力。对于开环 DNA 来说(有自由末端的 DNA),这种张力以两条链相互间的旋转来衡量,而对于共价闭环 DNA (环状 DNA 或两个末端固定的 DNA)来说,由于不能反向旋转,它的扭转张力必须以超螺旋的情况来衡量。

超螺旋的量度　　特定 DNA 的超螺旋程度可用**超螺旋密度**(superhelical density, λ)来表示。其计算公式如下:

$$\lambda = \frac{\tau}{L_0} = \frac{(L - L_0)}{L_0}$$

超螺旋绕数(superhelical winding number, τ)是指 L 与 L_0 间的差值。处于过度缠绕状态的 DNA 称为正超螺旋,τ 为正值;处于缠绕不足状态的 DNA 称为负超螺旋,τ 为负值。τ 值定量描述了分子所承受的扭转张力及形成超螺旋的倾向,但它并不表示确切的超螺旋数,因为螺旋的螺距会因扭转张力而改变。超螺旋数可用**缠绕数** (writhing number, W)来表示,它与连环数 L 的关系是:$L = T + W$,其中 T 是扭转数 (twisting number),它指 DNA 分子中总的螺旋数。连环数是 DNA 分子的拓扑性质(在变构时它不会改变),所以任何引起 W 值的改变都伴有 T 值等值的负向变化。在一个松弛分子中,$L = T$,因而 $W = 0$,所有的缠绕都是由螺旋引起的。一个单位的缠绕相当于 0.5 个单位的超螺旋,即相对于 DNA 螺旋轴旋转 180°。当一个超螺旋分子被置于水平面上时,两个双螺旋交叉之处被称为节(node),这个点可被认为是一个缠绕单位。

链状排列及结　　同闭环分子的螺旋缠绕特性一样,**链状排列** (catenation, 环状 DNA 的内部交锁)及形成**结**(knots)都是 DNA 的拓扑性质,只有打开闭环 DNA 分子才能破坏这种结构。当 DNA 分子被置于一平面上时,这两种结构都会出现由于两条双螺旋交叉而形成的节。所以,对给定的 DNA 分子来说,总的连环数可用连环数加上连锁与结的数目来表示,即总链环数 = $L + C + K$。

(殷向雷 译)

第 17 章
核酸结合蛋白

基本概念和定义

- 同核酸相互作用的蛋白特别重要，因为它们控制一些非常基本的生命过程，这些过程包括 DNA 复制，重组，DNA 的修复，转录，RNA 的加工和蛋白合成。尽管蛋白核酸的相互作用在细胞中的功能多样，但是它们的功能主要可分为四类：(1)结构和包装作用：(例如：染色体中的组蛋白，细菌类核(bacterial nucleoid)的 HU，病毒的衣壳蛋白) ；(2)运输和定位作用：包括 DNA 在核仁中分离和定位，RNA 的输出和定位，质粒的转移；(3)代谢和重排作用：(例如：DNA 和 RNA 聚合酶，核酸酶，解旋酶，DNA 修复酶，重组酶，拓扑异构酶)；(4)基因表达作用：例如：(RNA 聚合酶，转录因子，核糖体和起始因子，RNA 剪接装置，氨基酰 tRNA 合成酶)；对于某种蛋白，这些功能中很多是重叠的(例如，组蛋白是包装蛋白，但是又可影响基因表达，又如 RNA 聚合酶既可完成基因表达又涉及到 RNA 的代谢)。所有的核酸在它们生活史中的某个时期需同蛋白作用，并且许多核酸以**核酸蛋白复合体**的形式存在。
- DNA 通常以双股长链的形式存在，并且有相对均一的螺旋结构，而 RNA 主要以单链形式存在，并且其二级和三级结构形式多样。这些差别同时也反映在 DNA 和 RNA 同蛋白相互作用的原理和复杂性上。
- 核酸结合蛋白可根据其底物的特殊性分为三大类：①非特异性结合蛋白；②序列特异性结合蛋白；③结合非正常结构的蛋白(例如在 DNA 中的不正常碱基、重组中间体等以及 RNA 中剪接套索)。相互作用的模型可能涉及末端基团的结合，或将核酸封闭在窄缝或环中，大部分核酸结合蛋白在一个内部位点以定位方式相互作用。常常涉及用作识别因子的蛋白中的特定的二级结构，并且往往根据包含这些因子组件的结构，将核酸结合蛋白分成不同的家族。对于 DNA 结合蛋白，这些组件常常包含一个穿越 DNA 大沟的 α－螺旋。与此相反，β－折叠和其他一些具有平直的表面的二级结构在 RNA 结合蛋白中出现的频率更高。
- 蛋白可以同核苷酸的主链或碱基相互作用，在双链分子中，需在大沟和小沟处结合。许多不同类型的非共价键被用于蛋白同核酸的相互识别中，主要包括静电吸引和氢键，主链连接、氢键和碱基连接。氢键常常涉及界面上有序的水分子的使用。蛋白和核酸相互作用通常通过两种分子的构象改变来描述，这样可产生最大的互补作用表面，其中包括一些包埋的接触面；范德华力对于相互作用十分重要。序列特异性结合可能是由碱基所显示的不同结合模式识别来完成，或者是识别

主链的不同构象。许多序列特异的DNA结合蛋白以二聚体的形式存在，这样增加了核酸蛋白相互作用的灵敏性和特异性，并且产生协同结合的效应。核酸结合蛋白的二聚化还可以用于增加识别的多样性和可调控的程度。

17.1 核酸被蛋白识别

DNA识别的一般方式 胞内DNA主要以很长的双链分子(dsDNA)形式存在，同RNA不同，双链的总体结构相对稳定(参见双链螺旋)。DNA结合蛋白识别整个螺旋分子的几何结构，通过接触大沟或小沟(参阅)，连接不变的核糖磷酸主链或碱基同DNA相互作用。然而，螺旋的构象是多种多样的，这些不同不仅可由于外部环境造成（例如氢键)，还可能是局部碱基堆集的相互作用造成，这种作用为顺序特异的。局部结构的多形性不仅影响了碱基同螺旋轴之间的关系(见文框6.1对螺旋参数的讨论)，而且还影响了螺旋的周期性和大沟，小沟的大小。碱基序列也决定了DNA内在的可弯曲的程度，局部的变化可通过决定在蛋白结合时成键原子间的空间组成和DNA可改变的能力大小，从而控制了蛋白和DNA之间的空间化学关系。

同双链DNA(dsDNA)结合的蛋白可被分为三类：①同DNA末端相互作用的蛋白(如DNA连接酶，外切酶)；②围绕DNA或以深的狭缝结合DNA的蛋白(如DNA聚合酶，拓扑异构酶)；③同DNA双螺旋的表面作用的蛋白。前两类酶包括了许多DNA加工的酶，而后一类成员最多，包括大多数转录因子，限制性内切酶，DNA包装蛋白，位点特异的重组酶和DNA修复酶。因而，后一类蛋白不仅包括普通的和序列特异的DNA结合蛋白，而且还包括识别非正常核酸结构(如损伤碱基等)的蛋白。还有一些蛋白同单链DNA(ss-DNA)作用如RecA(参阅)，SSB(参阅)，丝状噬菌体M13的衣壳蛋白。

那些同双螺旋表面相互作用的蛋白通常具有一种十分符合大沟的基序。这种结合通过形成稳定结合的埋藏接触面使接触区最大化。侵入大沟对于序列特异的蛋白结合是十分重要的，这样有助于通过直接结合碱基进行序列识别。对于这些相互作用来讲，结合大沟比结合小沟更适合，因为它更大(可容纳α螺旋，β带结构，折叠链和环状结构)并且键形式不固定。相反，在小沟处，不同碱基对产生同样的成键形式（见下文)。同磷酸主链的结合允许普遍性识别，可以稳定同大沟结合的结构，不过，由于局部碱基顺序可能会影响主链的三级结构，因而同主链结合也用于序列特异结合。

根据用于DNA识别的结构基序不同(表17.1，和下面章节)，将DNA结合蛋白分为不同的家族，通常这个基序中包含着一个侵入大沟的α螺旋，但是在另一些家族中，一个β带结构或环状结构代替了α螺旋的作用，相互作用涉及到互补表面，结合后可能会改变构象以使蛋白和核酸的特异化学基团相互作用。虽然由于碱基堆集而成的静电键，范德华力和分散力对于整个识别和稳定结构是十分重要的，但是从识别的特异性来讲，氢键是主要形式。

表 17.1 DNA 和 RNA 结合蛋白中基本的识别结构。这些结构定义了下文中讨论的蛋白家族。最大的家族(HTH、锌指和碱性结构域家族)可以进一步分为不同的亚家族

结构	识别	分类学分布	举例
DNA-结合结构			
螺旋-转角-螺旋	α-螺旋	所有的	大肠杆菌 Lac 阻遏物,果蝇的触角足蛋白
HMG 结构域	α-螺旋	真核生物	大鼠的 HMG1
锌指结构	α-螺旋	主要在真核生物	小鼠的 Zif268,果蝇的 Tramtrack
类固醇受体家族	α-螺旋	真核生物	人类维甲酸受体,大鼠糖皮质激素受体
双核基因簇	α-螺旋	酵母	酵母 GAL4
碱性结构域	α-螺旋	真核生物	酵母 GCN4,人类 MyoD,小鼠 Max
BPV E2 基序	α-螺旋	乳头瘤病毒	BPV E2 蛋白
带-螺旋-螺旋和其他片层结构	β-折叠	所有的	大肠杆菌 MetJ 阻遏物,拟南芥 TATA 结合蛋白
组蛋白 —核心	静电作用	真核生物	人类 $TAF_{II}31$,$TAF_{II}80$
—接头	α-螺旋		人类 H5
Rel 同源蛋白	环	真核生物	NF-κB p50 同源二聚体
RNA-结合结构			
RNP 结构域	β-折叠	真核生物	U1A snRNP
dsRBD	β-折叠	所有的	果蝇的 Staufen
K-同源蛋白	环	真核生物	hnRNP K

缩写:HMG=高泳动蛋白;TAF=TBP(TATA 结合蛋白)相关因子;BPV=牛乳头状瘤病毒;RNP=核糖核酸蛋白;dsRBD=双链 RNA 结合蛋白;hnRNP=核内不均一核糖核蛋白。

RNA 识别的一般方式 不同于 DNA,胞内 RNA 通常以单链形式存在,但往往折叠成各种二级结构(突起,发夹,茎环),这些结构可作为蛋白结合位点,并且还能形成更为复杂的三级结构(三股螺旋,假结)。RNA 这种可形成高级结构的能力使其具有更大的构象多样性,使 RNA 还具有了同蛋白质一样的催化功能(参见核酶),这些特点意味着 RNA-蛋白相互作用的机理将更为复杂。

如同 DNA 一样,某些 RNA 结合酶特异性地同末端基团作用,并将一些底物封闭入隧道或狭缝中。然而大部分 RNA 结合蛋白以一种局部形成的方式同 RNA 进行普遍性或序列特异性的相互作用。尽管对于 RNA-蛋白作用的潜在复杂性,从许多蛋白 RNA 结构的最新解释已经确立了一些共同的主题(见表 17.1 和下面章节)。值得特别注意的是,β 折叠经常用于相互识别的表面,可能是由于这种结构使暴露的 RNA 碱基利于伸展而更合适化学结合,当在 RNA 分子内出现双链 RNA 时,2′-羟基减小了双螺旋的柔性,这种双螺旋采用了 A-DNA 的典型构象(参阅),在这种结构中,小沟宽而浅,而大沟窄而深,在一些 RNA 病毒的衣壳蛋白中,就发现它同小沟作用。但是,由于 RNA 环化或弯曲而导致扭曲,可以扩大大沟,这样一些典型的 DNA 结合结构,如同源结构域也可以同 RNA 结合。

17.2 蛋白中的DNA结合基序

螺旋－转角－螺旋基序　**螺旋－转角－螺旋(HTH)**基序(motif)是第一个被确立而被广泛研究的DNA－结合结构。在大肠杆菌和其噬菌体中,许多经深入研究过的基因调控蛋白中发现了经典的HTH。蛋白－DNA复合物结构经X光衍射结晶和NMR方法解析成功:Lac阻遏蛋白，Trp阻遏蛋白、分解代谢激活蛋白(CAP)，λ阻遏蛋白λ－Cro，噬菌体434阻遏蛋白，噬菌体434 Cro PurR(参见噬菌体λ、lac操纵子、转录调节－细菌).还有更多的含有HTH结构的蛋白在不含DNA的情况下解析成功。

这种基序包含有两个α螺旋,螺旋之间间隔有一个短的β转角,使两个螺旋可通过疏水作用装配起来(图17.1)。第一个螺旋稳定并使第二个螺旋暴露出来,与DNA的大沟作用,而特异性地与碱基接触。因而,尽管两个螺旋,或包含HTH的结构域中的其他残基也可能参与了同DNA的相互作用,但第二个螺旋仍被称为**识别螺旋**(recognition helix)。上述的相互作用锚定了蛋白质中识别螺旋的位置并稳定了DNA的构象,从而调节了不同蛋白和其结合位点的亲和力。

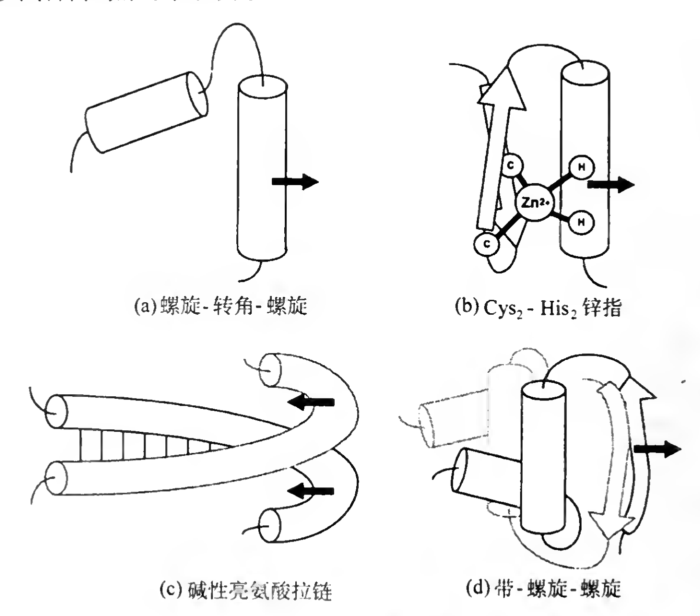


图17.1　共同的DNA结合基序。(a)从λ噬菌体来源的螺旋－转角－螺旋基序。(b)从爪蟾Xfin来源的 Cys_2His_2 锌指组件。(c)从小鼠Max来源的亮氨酸拉链/螺旋－环－螺旋二聚体基序的α螺旋(线代表了在卷曲的二聚体界面中的亮氨酸之间的键)。(d)从噬菌体P22 Arc阻遏物来源的带－螺旋－螺旋基序。箭头代表识别结构,侵入大沟与碱基形成氢键。二聚体以阴影表示。

典型的细菌HTH蛋白有一个高度保守的转角结构,包含有4个残基,其中第二个残基往往为甘氨酸,在另外一些非保守的细菌HTH蛋白中[如,LexA(参阅)和AraC],这种限制是不严格的。真核细胞的HTH蛋白有着变化更大的转角和环结构,依据上述特点,总体结构和序列保守性,可将蛋白分为许多家族。最大和研究最仔细的是**同源结构域**

(homeodomain)家族,同源结构域首先从果蝇中的触角足蛋白(Anternapidia proteins)中被发现,而且这种结构在许多控制内环境稳定(homeostasis)的转录调控因子和发育过程中的决定区域特异性的蛋白中占主导地位(参见同源异形基因)。在 DNA 修复和染色体结构的蛋白中,同源结构域蛋白也有发现。同源结构域是一个 60 个氨基酸的保守结构域,含有四个 α 螺旋,螺旋Ⅱ和螺旋Ⅲ之间由 β 转角相连并成直角,左右形成了 HTH 基序,螺旋Ⅲ作为识别螺旋。

与含有 HTH 结构密切相关的结构包括 POU 特异结构域和成对结构域。POU 转录因子包含两个 HTH 基序,一个在同源结构域中,另一个形成 **POU 特异结构域**,它同 λ 阻遏蛋白的 HTH 结构十分类似。同源结构域和 POU 特异性结构域之间有螺旋间隔的连接臂,POU 转录因子可同时结合 DNA 双螺旋的对侧。两个结构域中的任一个都可单独结合 DNA,但是同时结合增加了结合的特异性和稳定性。由于 POU 特异性结构域首先在脊椎动物的转录因子——*P*it-1, *O*ct-1 和 *O*ct-2,以及线虫的 *U*nc-86 蛋白中发现,因而用这些基因的首写字母命名。

同源结构域蛋白中的配对家族(包括果蝇的配对和以及很大脊椎动物的 Pax——配对盒——转录因子可能也含有两个 HTH 基序。配对相关蛋白含有第二个 HTH 基序和一个与 Hin 重组酶同源的组件,一道组成了独特的**配对功能域**(paried domain)。

其他的真核细胞的 HTH 蛋白还包括同酵母热休克调控因子有关的一大家族以及被称为翅状螺旋(winged-helix)转录因子的 HNF3/Fork head 的家族。**高泳动蛋白功能域**[high mobility group(HMG) domain]也包含着由 β 转角连接的系列 α 螺旋的结构,它存在于染色体结构蛋白和如睾丸决定因子 Sry 等转录因子中,也由转角隔开的几个 α 螺旋组成。然而折叠的几何学与典型的 HTH 基序不同。HMG 蛋白可以和 dsDNA 的小沟作用。因此它们属于与 HTH 蛋白不同的 DNA 结合蛋白家族。

细菌和真核细胞的 HTH 基序在拓扑上近似,第二个螺旋都作为识别螺旋,并且蛋白中的其他部分帮助识别螺旋同 DNA 的大沟作用(同时同 DNA 产生其他的相互作用)。但是,从识别螺旋在大沟中的方向来看是不同的。细菌的识别螺旋的 N 端区域对于结合是至关重要的,同时由同源结构域识别螺旋的中部区域完成其结合功能。而且,虽然细菌识别螺旋的方向变化大 (如在 λ-阻遏蛋白中;其顺着大沟排列。而在 Trp 阻遏物中,识别螺旋又与大沟垂直),而同源结构域识别螺旋的方向是十分保守的。细菌的 HTH 蛋白普遍以同二聚体的方式作用,识别 DNA 的回文序列。直至最近,含有同源结构域的蛋白还被认为在不对称的位点以单体的形式作用。但是,有越来越多的证据表明同源结构域在体外可以二聚化,几种同源结构域的同二聚体同 DNA 结合的复合物和异二聚体同 DNA 结合的复合物的结构被解析出来(例如果蝇的配对同二聚体,酵母交配型调节因子 MATα1 和 MATα2 的异二聚体)。同源结构域蛋白也可以同非同源结构域蛋白形成二聚体(例如,MATα2 同 MCM1 二聚化,而且许多果蝇同源结构域的辅助因子,如 Teashirt 已经被鉴定)。

半胱氨酸-组氨酸锌指　**锌指结构**(zinc finger)是第一个被发现的真核细胞中同 DNA 结合有关的结构,并且也是已知存在最广泛的蛋白组件之一,大约占真核基因组的编码序列的 0.5%(参见蛋白质家族)。含锌的蛋白在结构上多样,并且可根据锌配位的机制和结构域结构将其分为至少 6 个大家族(图 17.2,另见后面的章节)。相对来讲,锌指结构

在细菌中十分少见：大肠杆菌 DNA 修复蛋白 MutM 是具有含锌配位组件的细菌蛋白的少数例子之一。

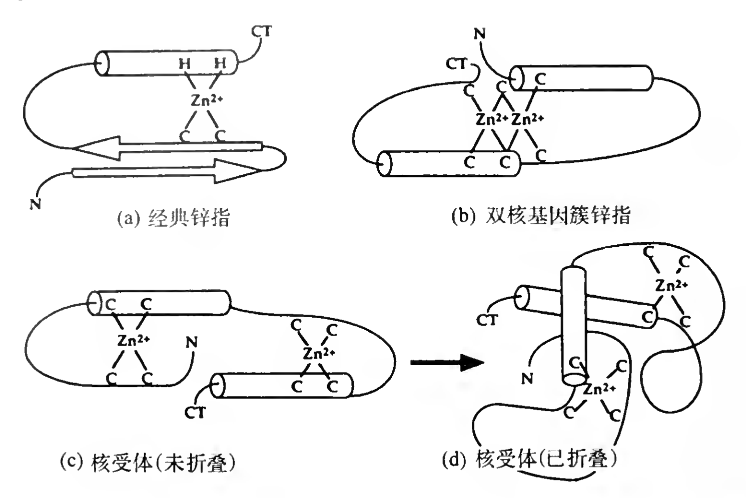


图 17.2 三种主要类型的锌指等同结构的结构。(a)经典的 C_2H_2 锌指。(b)双核基因簇 GAL4 的 C_6。(c)核受体的 $2\times C_4$ 锌指，没有折叠以显示拓扑学结构(转录因子 GATA－1 有类似结构的单个 C_4 锌指)和(d)折叠以显示结构域结构，N 末端识别螺旋。α 螺旋以圆柱表示，β 链以箭头表示。C＝半胱氨酸，H＝组氨酸，N＝N 末端，CT＝C 末端。

Cys_2His_2 锌指包含有一对半胱氨酸残基和一对组氨酸残基，这些残基在 12 个氨基酸残基组成的环中与锌离子配位。这个环状结构也包括一些保守的疏水残基和许多碱性残基，其核心的保守序列是

$$\mathbf{C}-X_{2-4}-\mathbf{C}-X_{2-4}-F-X_5-L-X_2-\mathbf{H}-X_{3-4}-\mathbf{H}$$

这个环状结构凸出于蛋白表面(因此称为手指)，其顶端的碱性残基同 DNA 分子的大沟作用。锌指采用的结构被认为是 β 发夹和一个相邻的 α 螺旋(包含有 His 残基)，前一种结构同 DNA 的主链作用并使螺旋维持在大沟处，作用产生序列特异性的接触(图 17.1)。除去锌后，或将保守的半胱氨酸突变，或阻止锌的配位，蛋白的功能都将丧失，这些说明锌对于正确构象的维持是十分重要的。

第一个锌指基序图样发现于爪蟾基本转录因子 TFIIIA 中。这个蛋白含有 9 个串联起来的锌指，并且由柔性的连接子连接起来的多个锌指是此家族 DNA 结合蛋白的特点之一。大鼠转录调控因子 SP1 和 Zif268 含有三个锌指，而果蝇的转录因子 Tramtrack，Kruppel 和 Hunchback 分别含有 2，4，6 个锌指。著名的爪蟾蛋白，xFin 含有不少于 37 个锌指！

几种锌指蛋白同 DNA 结合的复合物的结构已经解析得出。表明多个锌指沿着大沟螺旋方向同 DNA 碱基形成一系列的结合。每个锌指结构识别三个碱基一组的密码，这种识别由氢键介导(如 Zif268 在 9bp 的识别顺序中形成 11 个氢键)。至少对于类 Zif268 锌指结构，一个完整的氨基酸与 DNA 碱基识别的密码的简单规则已经出现(见下文)。对于一些锌指蛋白，虽然同 DNA 作用的整个几何形状是保守的，但不是每一个锌指都参

与同碱基的特异识别,许多锌指一般作为隔离物作用,它们根本不与 DNA 直接接触。

半胱氨酸锌指　　典型的锌指组件包含有一个锌,其由两个半胱氨酸和两个组氨酸残基配位。而其他许多含锌的 DNA 结合结构域只用半胱氨酸配位锌,因此,有时被称为**多半胱氨酸锌指**。

最大的多半胱氨酸锌指蛋白家族是转录因子中的类固醇胸腺激素**核受体家族**。每个这种蛋白包含有二个指头,每个指头中有 4 个半胱氨酸配位锌离子,并且每个指头与 α 螺旋 N 端的延伸部分共同形成(图 17.2)。这种锌指缺乏 Cys_2His_2 锌指具有的保守疏水残基,其参与形成包含着有垂直的 α 螺旋的一个单个的,球形的 DNA 结合结构(图 17.2)。N 端的螺旋是其识别螺旋,而第二个螺旋可能与其二聚化有关。类固醇受体以同二聚体的形式结合于 DNA 的回文靶位点。而胸腺激素,维生素 D 和视黄酸的受体主要同其他家族成员——视黄醛 X 受体——形成异二聚体,与正向重复部分结合。一些核受体能以单体与 DNA 结合,这种情况至少在体外实验中可观察到。

其他多半胱氨酸锌指还包括造血的转录因子 GATA-1,与酵母转录调节因子 GAL4 有关的一些蛋白及一种人的延伸因子。GATA-1 的结构非常类似于核受体家族,虽然它仅有单个的锌结合组件。相反地,GAL-4 相关蛋白的结构是独特的,至今只发现于酵母蛋白中:6 个半胱氨酸配位 2 个锌原子。这样中间的半胱氨酸参与 2 个锌离子配位(图 17.1)。有时,这种结构又被称为锌双核簇(zinc binuclear cluster),它具有下列保守序列:

$$C-X_2-C-X_6-C-X_6-C-X_2-C-X_6-C$$

这种的锌配位结构被认为帮助 α 螺旋存在于大沟处。但是,在人的延伸因子中,为一个 β 折叠同 DNA 结合。p53 肿瘤抑制因子却用一个 α 螺旋侵入 DNA 的大沟,而含锌配位结构使存在于小沟的残基呈稳定的环状结构。

碱性结合结构域　　真核细胞的转录因子中的一大类家族都含有一个高度碱性的 α 螺旋,作为同 DNA 的基本识别结构。在许多情况中。这个螺旋直接同二聚化结构域相连,或是一个**亮氨酸拉链**(leucine zipper)[因此又称为**碱性亮氨酸拉链**(basic leucine zipper, **bZIP**)]或是一个螺旋-环-螺旋结构[因此又称**碱性螺旋-环-螺旋**(basic lwliz-loop-helix, **bHLH**)],或者两者都有。这种结构形成螺旋卷曲,这样使碱性结构域维持在某一构象,允许它们可与 DNA 的对侧作用。它们同不同的家族成员形成同源或异源的二聚体使 DNA 结合家族更具功能的多样性,更易于调控。

识别螺旋可能是作为 DNA 结合的一种后果:酵母 bZIP 蛋白 GCN4 的研究表明,碱性结构域在溶液中采用一种无序的部分螺旋的结构,不过,当同 DNA 结合时,这种构象发生了变化从而诱导了典型 α 螺旋的形成。在 GCN4 的二聚体中,识别螺旋是刚性结构,其仅以几个碱基同 DNA 轻微接触。但是在小鼠的 Max 蛋白中,螺旋在中部弯折,围绕 DNA 折叠成剪刀柄式,紧贴于 DNA 的大沟,形成更广泛的接触(图 17.1)。突变分析表明:二聚化结构域同碱性结构域一样都是 DNA 结合所必要的。这是由于二聚化结构域中的残基参与了同 DNA 的结合,并且要形成同 DNA 高度亲和的结合,也需二聚化;如果人为通过二硫键连接一对单独的碱性结构域,它可与 DNA 结合。但是,一些碱性结构域蛋白缺乏拉链/HLH 二聚化基序(如果蝇中的蛋白 Mastermind)。另外,在一些 HTH 蛋

白(如酵母的热休克因子)和一些具有锌指结构的蛋白(如GAL4)中,拉链也作为二聚化的结构作用。

类组蛋白结合基序 真核细胞的DNA分子可通过同组蛋白作用被包装成重复的结构单位,称为核小体(见第3章染色质)。在核小体的核心颗粒中包含有二个负超螺旋弯曲的DNA围绕着一个组蛋白的八聚体。这个八聚体分别由同样两个核心组蛋白H2A,H2B及H3,H4组成。它们同DNA之间的相互作用涉及DNA中富含AT区域的小沟出现在组蛋白八聚体的表面并将其缠绕。一级序列排列(primary sequence alignments)揭示:各组蛋白之间仅有很低的同源性(<20%),但是它们具有一个保守的结构,即**组蛋白折叠**(histone fold),它包含着由二个介导二聚化的融合的螺旋-链-螺旋的结构,组蛋白的连接子(如H1,H5)也在染色体的高度有序化中起作用(参见 30 nm 纤丝),它们通过结合DNA进入和离开核颗粒的区域封闭了核小体。

许多转录因子与不同的核心组蛋白同源,其中包括同基本转录因子IID(TFIID)相关的TAFs(参阅)。例如,人类的$TAF_{II}31$同H3同源,而$TAF_{II}80$同H4同源。$TAF_{II}20$同H2B同源,但是缺少同H2A同源的基本因子。$TAF_{II}20$,$TAF_{II}31$和$TAF_{II}80$可以组成类似于组蛋白八聚体的四级结构,其中$TAF_{II}20$形成同二聚体;还有证据表面,TFIID诱导了DNA负超螺旋的形成。负调控因子DR1,可抑制TFIID的作用,并且这种作用可被TFIIA阻断(参见 转录起始——RNA 聚合酶 II),它同组蛋白H2A同源,DR1同一个称为DRAP1的蛋白形成二聚体产生作用,这个二聚体同H2A/H2B的二聚体类似。

连接子组蛋白H5已解析的结构中发现:在其球状结构域中,存在着一个螺旋-转角-螺旋的基序。正如上面讨论的那样,这种基序在许多真核和原核细胞复制起始区的转录因子中可找到。组蛋白的HTH基序包含于一个大的结构域中,这个结构域还包括其他的螺旋,链和环。在属于HNF3/Fork head家族的真核HTH转录因子的翅状螺旋结构域中和大肠杆菌中,*bio*操纵子的阻遏物BirA中可以找到相似的拓扑构象。

乳头瘤病毒的E2蛋白 这种蛋白具有一种独特的结构,包括4个β链和一个突出的α螺旋。4条链形成一个弯曲的β-折叠,可作为二聚化的作用表面。因此,整个二聚体蛋白具有一个穹顶状,由8条链组成的β-桶状结构,其中还带有二个突出的α螺旋。DNA围绕此蛋白规则地弯曲,α螺旋在两个连续的大沟中形成碱基特异性识别。

通过β-链识别 在上述讨论过的大部分蛋白中,与DNA的识别序列中涉及一个α螺旋,存在于大沟。然而,在三种细菌阻遏物(MetJ,Arc和Mnt)中发现了另一种称为**带-螺旋-螺旋**(ribbon-helix-helix)的基序,它同DNA的识别是由一个β折叠介导的(这个带是一二条反向平行β折叠链组成适合于进入DNA的大沟)。这三个阻遏物以二聚体的形式作用,每个单体提供一个由二个α螺旋稳定的单个β链。通过四级结构的相互作用,β链呈带状排列,顺大沟放置,通过形成多个氢键有利于识别(图17.1)。二聚化对于构成DNA结合结构域是必需的。细菌类核包装蛋白HU可能也采用相类似的作用模型。

在真核细胞中,负责真核基因基本转录的 TATA 结合蛋白(TBP)(参阅)也可以通过一个反平行的β-折叠同DNA作用。这个蛋白明显为一个单链多肽,但是它形成了两个结构非常相似的结构域,这样形成一个对称的C字形分子,象马鞍状包围DNA分子。在

弧形表面的β-链的残基是同DNA作用的主要残基，而在弧形外表面的α螺旋同其他基本的转录因子作用。与细菌的绶带状阻遏物不同，TBP蛋白中的β链结合DNA的小沟。

The Rel 家族　　在真核细胞转录因子中的Rel家族对于各种不同的环境刺激产生效应，并且在发育过程中也起着十分重要的作用。Rel转录因子中包括果蝇的和哺乳动物的NF-κB。Rel蛋白有一个保守的结构，即**Rel同源结构域(RHD)**[1)]，它全长大约300个氨基酸残基，由一个柔性的铰链区连接二个免疫球蛋白折叠构成。虽然两个折叠中的残基都可以同DNA作用，但是它们的分工明确，其中的一个是二聚化的结合面，另一个是直接同序列特异DNA相互作用。Rel蛋白可以形成同二聚体和不同的异二聚体，都具有二聚化的倾向。同二聚体和异二聚体结合不同的DNA位点，但相互作用是相当混乱的。Rel蛋白结合的组成和二聚化的混杂性，以及细胞类型和发育的特异表达，为复杂的调控网络作出了贡献。一些Rel蛋白也可以以单体同DNA结合。

二体Rel蛋白包括原始型(anchetypal)的转录因子NF-κB，它由两个亚基——P50和P65(即RelA蛋白组成)。P50同二聚体同DNA结合的复合物的解析表明，它同DNA的相互作用主要由位于DNA大沟中的Rel的一个长的环状结构所介导。在这个环状结构中的残基同5个连续的DNA碱基结合，而更多的位于环状结构外的残基可以同DNA产生序列特异性结合。

DNA加工酶中的结合基序　　DNA加工酶同其底物结合的结构是多种多样的。那些催化合成和拓扑修饰的酶以其沟状或环状结构围绕DNA。第一个得到解析的此类基序来源于*E. coli*的DNA聚合酶I的Klenow片段，这个酶有一个深沟，首先此深沟同DNA结合，随后再围绕DNA折叠并将DNA包裹起来。但对于大肠杆菌中的DNA聚合酶III却有着不同的策略：其β亚基是由新月状亚基组成的二聚体，像一个滑动的夹子，围绕DNA形成一个环。真核细胞中的PCNA(参阅)也以同样的方式作用，但是它是三聚体。裂缝结构在其他一些聚合酶中也有发现：例如，HIV的反转录酶，RNA聚合酶II，它们都被认为用同样的方式同其底物作用。HIV反转录酶与核酸复合物的高分辨率结构表明：结合是通过一个包埋于DNA大沟中的酶的一个环状区域同DNA磷酸主链作用。

功能简单的酶，如核酸内切酶，同DNA螺旋的表面结合，这种蛋白家族我们在前面已有所讨论。此类酶的多数采用侵入大沟或小沟的结构，但是它们比转录因子同DNA的相互作用更为广泛，它们的结合结构域部分地包裹螺旋。几种限制性内切酶同DNA的结合复合物的结构已经被解析，其中揭示出许多不同的DNA结合基序。通常的观点认为限制性内切酶倾向于使DNA发生剧烈扭曲。*EcoR* I诱导一个尖的纽结，将酶切位点处的大沟或小沟扩大，这样可以允许一个大的四个桶状的α螺旋置于大沟之中，而延伸处的手臂将围绕这个螺旋。相反，*Eco*RV利用环状结构同靶位点的大沟作用，非特异性的DNase I也是如此。

17.3　蛋白中RNA结合基序

RNP结构域　　最为了解的RNA结合结构域，即核糖核蛋白(RNP)结构域，首先在酵母

---

1)原文误为RHB，应为RHD(Rel homology domain)，译为Rel同源结构域。——译者注

的多聚腺苷结合蛋白(polyadenylate-binding protein)(参阅)中发现。此类结构也有两种保守基序——RNP1 和 RNP2,它们都对 RNA 特异性结合十分重要,而且它们仅是一个更大结构的一部分,这个大的结构由 4 个 β 链和 2 个 α 螺旋以 β-α-β-β-α-β 的拓扑结构组成。RNP1 和 RNP2 基序位于一个四条 β 折叠的中心 β 链处。

U1A 剪接体蛋白含有二个这样的结构域,它可以结合 U1 SnRNA 的一个发夹结构,这种 RNA 蛋白复合物的结构已被解析,它由 β 折叠形成一个非特异性的 RNA-蛋白结合平台,U1 SnRNA 的发夹结构和一个富含负电荷残基的环状结构之间的相互作用,在折叠稳定了 RNA 的构象中连接两个 β 链,这样 RNA 可以特异地同 RNP1 和 RNP2 作用。

双链 RNA-结合结构域　双链 RNA 结合结构域(dsRBD)是一个短的(大约 65 个残基)的组件,它存在于好几种可结合双链 RNA 的蛋白中(如:腺苷脱氨酶, dsRNA-依赖性激酶,果蝇 Staufen 和大肠杆菌 RNase III)。它们的拓扑结构类似于 RNP 结构域(形成一个 β 折叠的三个 β 链和两个 α 螺旋,带有 β-α-β-β-α-β 的拓扑排列),但是同 RNA 的特异性结合仅发生于此结构域中第二个 α 螺旋和 β 折叠形成的裂缝中。

K 同源结构域　K 同源结构域首先在核不均一 RNA 结合蛋白 hnRNP K 中发现,这个蛋白含有三个这种结构域。在几种其他的 RNA 结合蛋白,其中包括脆性-X 基因 *FMR*1 的产物中也发现了这种结构。同 dsRBD 一样,K 同源结构域以三个链组成的 β 折叠为中心,但是它们还具有三个 α 螺旋(拓扑组成为 β-α-α-β-β-α),特异性相互作用被认为产生于 RNA 和头两个 α 螺旋连接区的残基之间。

其他 RNA 结合结构　在上面讨论过的一些真核细胞的 DNA 结合蛋白同样也可以结合 RNA。锌指蛋白 TFIIIA 可同 5S rRNA 作用作为调控蛋白合成的一种形式,而果蝇的同源结构域蛋白 Bicoid 可以通过结合 *caudal* mRNA 而阻遏 Caudal 蛋白的合成。由于 RNA 的大沟太窄,从而不能容纳这些结合蛋白常规的识别螺旋,因此,很可能是由不同的残基参与同 DNA 和 RNA 的结合,或者是识别位点的 RNA 产生严重扭曲,这样可以允许其侵入大沟。最新解析而得的核糖体蛋白 L11 的结构显示其具有与同源结构域结构类似的 HTH RNA 结合基序,而 HIV Rev 中被认为是通过一个 α 螺旋在一个加宽的大沟处结合。这个 α 螺旋也是 ColE1 质粒编码的 Rom 蛋白作用的基本识别基序(见第 20 章质粒)。然而在这种情况中,一个四螺旋束以 β 折叠相同的作用方式提供一个平面来展示和解读 RNA 碱基。相反地,在 HIV 的 Tat 蛋白中,由一个 β 带侵入 *Tar* 大沟一侧,以利于进行碱基的识别。许多核酸结合结构偶尔也用于蛋白与蛋白的相互作用。

17.4　蛋白,核酸结合的分子基础

核酸与蛋白的直接作用　核酸蛋白直接作用的四种途径为:①蛋白侧链同碱基作用(是序列特异性相互作用的主要形式);②蛋白主链(如氨基基团)同碱基作用;③蛋白侧链同磷酸骨架的作用;④蛋白主链同磷酸骨架的作用。

许多不同类型的非共价相互作用存在于蛋白和核酸的相互作用中:氢键,范德华力,疏水作用,球形吸引和排斥的静电作用以及特殊的静电键。此外,核酸自身的氢键,碱基堆集作用以及蛋白氨基酸残基侧链间的吸引和排斥力也参与其中。

蛋白和核酸磷酸骨架的相互作用常常涉及带负电的磷酸残基和带正电的氨基酸残基(如 Lys, Arg)之间的静电相互作用。此外,在核酸骨架和蛋白侧链或主链氨基之间可形成氢键。除了蛋白和核酸骨架的直接作用之外,远程的静电吸引在蛋白核酸相互作用中起着十分重要的作用。实验证实,将中性的氨基酸残基换成碱性残基,一些转录调控因子同核酸的亲和力可增强,甚至是当这些残基离核酸太远而无法直接同 DNA 作用的时候。与主链的接触也常常存在于非特异性结合之中,但是特殊的碱基序列可能会影响核酸骨架的三级结构,这种结构可以被蛋白识别和解读,产生序列特异性的相互作用。

蛋白和核酸碱基之间的相互连接识别,不管是开放性地(同单链),还是通过大小沟(双链分子),主要由碱基和氨基酸残基侧链之间形成的氢键介导。这种接触可能是一一对应的,但是几个侧链可以同一个碱基对连接,长的侧链还可以横跨好几个碱基对。序列特异性的核酸结合蛋白倾向于利用其埋藏表面,例如一个插入大沟的 α 螺旋,使它们之间的接触最大化。蛋白残基和 DNA 碱基之间的相互作用对于序列特异识别是十分重要的,但不是所有的接触都是特异的,例如,在糖皮质激素受体和 DNA 的相互作用时,就产生非特异性识别而形成的广泛的碱基和蛋白侧链的接触。

水在核酸蛋白相互作用中的作用　有序的水分子存在于核酸蛋白作用的接触面,它们或仅是接触空间的填充物,或者本身参与了键的形成。在非特异性结合蛋白中,水被认为是绝缘子,衬于接触的内表面,这样使得蛋白可滑动扫描并促进非特异性的相互作用。水也参与蛋白和 DNA, RNA 主链或碱基的氢键形成,在一些情况中,水也是碱基识别所必需的,例如,在 Trp－阻遏物－DNA 的复合物中,和在 $tRNA^{Gln}$同 Gln-tRNA 合成酶的复合物中。现在认为,水负责结合反应的亲和力和特异性。当 *Bam*H Ⅰ同其酶切位点作用时,四个有序的水分子介导蛋白侧链和碱基的接触,水也可能延伸由核酸蛋白组成的氢键网络,和增强识别的特异性。在许多的蛋白核酸的复合物中(例如锌指蛋白 Tramtrack 同其靶位点结合)这些水分子固定为一种特殊的构象,其极性表面也是刚性的。在其他的例子中(如同源结构域蛋白触角足结合其靶位点时)水分子可跳动,这样允许氨基酸残基侧链采取不同的构象,并在两个或多个结合态之间选择。结合面的流动性是很重要的,这是因为流动性增加了在接触面突变时(蛋白本身或其靶位点)蛋白结合的适应性;还可能是由于接触面的部分变形,增加了此系统的熵。水在蛋白同核酸骨架或大沟接触中起着十分重要的作用,而在小沟处的相互作用(如包括 HMG 蛋白,或者是基本转录因子 TBP)主要是由于疏水性,水好像被排出在接触面外。

核酸结构的蛋白调节　在 DNA 中特殊的碱基序列可以促进内部的暗带(anisotropic bending)(参阅)和呼吸(参阅),但是 DNA 在同蛋白作用时也发生了扭曲变形。蛋白可通过熔解碱基对,减小或增大弯折或螺旋来改变 DNA 的结构。许多 DNA 结合蛋白造成了 DNA 的一些结构变化,这些构象变化可形成两个接触面之间更佳的结合。DNA 的构象变化不仅对于最初接触面的调整起作用,而且还可控制选择性的结合和/或序列特异性 DNA 结合蛋白的活性。

最剧烈的扭曲和纽结是由限制性内切酶造成的,DNA 解旋对于加宽和暴露 DNA 的大沟催化活性中心是十分必要的。*Eco*R Ⅰ高度选择性地结合 DNA 中的限制性酶切位点,而 *Eco*RV 以同样亲和力结合于许多 DNA 序列。这是由于,对于 *Eco*R Ⅰ而言,底物位

点的选择性是由于在接触阶段对于序列特异的碱基的识别，而 *Eco*RV 对于底物的选择发生于变化阶段，这时仅有某些序列被解旋和纽结。转录因子也可以利用 DNA 的柔性。在转录复合体或增强体(参阅)处的 DNA 环便是一例，这样可引导适当的调控蛋白参与接触，而且其中常常涉及前面提及的 HMG 家族中的 DNA 弯曲蛋白。DNA 内源柔性在染色质结构的形成中也起着十分重要的作用(参见核小体，核小体分相)。

二聚化和协同性　许多转录因子，限制性内切酶和其他许多序列特异性的 DNA 结合蛋白以二体或更高聚合物的形式作用，在一些情况下，这是同 DNA 结合所必需的(如在带 - 螺旋 - 螺旋的阻遏物中)；而在其他的情况中，每个单体都有一个相对独立的识别结构，但是二聚化可以增加结合的特异性和灵敏度。综合起来可能有以下几种目的：①通过增加结合蛋白的大小，增加接触面和碱基识别的数量；②通过增加蛋白同 DNA 结合的接触面，增加结合的稳定性；③通过可循环的蛋白与蛋白作用，增加作用的亲和力；这种协同性的结合在许多转录因子中可以看到，包括那些甾类受体家族(在这里它们可通过改变重复识别序列的间隔被协同性地调节)。二聚化既可以通过异二聚体来增加识别靶位点的范围，又可以在调控中使用非活性单体来增加蛋白作用的多样性。

虽然蛋白相互作用的形式多样，但是许多转录因子还以二聚体形式作用。一些蛋白(如细菌 HTH 蛋白，类固醇激素受体)以同源二聚体的形式识别 DNA 中的回文序列。其他一些蛋白(如 bZIP 蛋白)可以在同家族中选择性地形成同源或异源的二聚体。同源结构域蛋白也可以在同家族的蛋白中选择性地形成同源或异源二聚体，还可以同其他家族蛋白形成异源二聚体。甲状腺素受体和其相关转录因子主要同一个简单分子形成异源二聚体，直接识别重复序列。还有其他一些蛋白以单体形式作用，但是它们内部有重复结构(如 POU 结构域转录因子包含二个 HTH 基序，可结合双螺旋的两侧，而 TATA 结合蛋白包含二个同样的 DNA 结合组件，锌指蛋白包括从两个到差不多 40 个的结合组件)。

二聚化对于 bZIP 和 bHLH 蛋白家族成员的作用是十分重要的。bZIP 蛋白通过**亮氨酸拉链**基序二聚化，亮氨酸拉链是一种两亲性 α 螺旋，可形成螺旋卷曲(参阅)。在二聚体中，每个螺旋的周期从 3.6 减至 3.5，这样可使亮氨酸每隔 7 个氨基酸产生相互作用。但是，在最初的模型中，亮氨酸残基被认为是交错对插，类似于拉链的齿，但现在认为这是不正确的，现在认为亮氨酸残基像梯子上的横档一样形成直接接触(图 17.1)。bHLH 蛋白有一个二聚化的螺旋 - 环 - 螺旋的基序，它同亮氨酸拉链相似，而不同于上面所讨论的**螺旋 - 转角 - 螺旋**的结构。一些蛋白同时具有拉链和 HLH 的二聚化结构，如 Max。

bZIP 和 bHLH 家族成员形象地诠释了功能和调节的多样性可由选择性地由二聚化产生。两种 bZIP 蛋白 c - Fos 和 c - Jun 形成被称为 AP - 1 的异二聚体转录因子；而 c - Jun 还可以形成同源二聚体，但是 c - Fos 不可能。肌浆中的 bHLH 蛋白 MyoD1，成肌素(myogenin)，Myf - 5 和 MRF - 4 的活性形式主要是异二聚体。然而，每种蛋白都同一种称为 Id(差异抑制物(Inhibitor of differentiation))的小蛋白二聚化，这种蛋白缺乏碱性结构域，可以阻止同 DNA 的结合。因而，Id 蛋白如同一个天然的显性失活突变体(dominant negative mutant，参阅)，通过使所有的活性 bHLH 蛋白形成非活性异二聚体的方法调节活性。

17.5 序列特异性结合

蛋白识别的 DNA 序列 许多序列特异性识别蛋白同 DNA 的复合物的结构已经被解析,其中一些蛋白造成了一个或多个碱基对的解离,但大部分识别发生在闭合双链的前后。总体上,相互作用可归为两种形式。**直接解读**(direct readout),它涉及在大沟或小沟处蛋白和碱基之间的相互作用:由于不是某一个键就可以明确地识别一对碱基,因此直接解读涉及接触的多个层面,其中有蛋白和 DNA 之间的作用,还有同作为桥分子的水分子的作用。**非直接解读**:涉及蛋白和糖,磷酸骨架之间的相互作用。在这种情况中,由于一些序列造成的整个 DNA 的特殊构象,蛋白可以识别这些序列。在 DNA 结合蛋白的结构分析中出现的一个最让人惊奇的发现便是在序列特异识别中非直接解读的突出地位。

DNA 中的分子信号 通过直接解读的序列特异性识别涉及蛋白和碱基间的键的形成。这表明:DNA 碱基中有一种隐含在键模式中的分子信号。这种信号可以通过研究在大沟和小沟处的化学基团加以确定,在这些位置,碱基暴露在溶液中。在大沟处,成键基团模式对于每个碱基对是唯一的,因而可在大沟进行明确地序列解读(图 17.3)。而在小沟处,成键模式是简并的,仅存在 A/T 和 G/C 之间的区别。在大沟处,某些键的位置也是不确定的,因此大,小沟中都可识别变化的序列。在一些转录因子,限制性内切酶等结合蛋白中,其识别位点是严格的不可改变,而对于其他一些蛋白,它们可以允许在一定位置上的某种程度的简并性,例如,在嘌呤和嘧啶之间,在酮基和氨基之间或在强和弱的成键碱基之间 (见图的说明,表 16.2 有关的解释和碱基符号)。

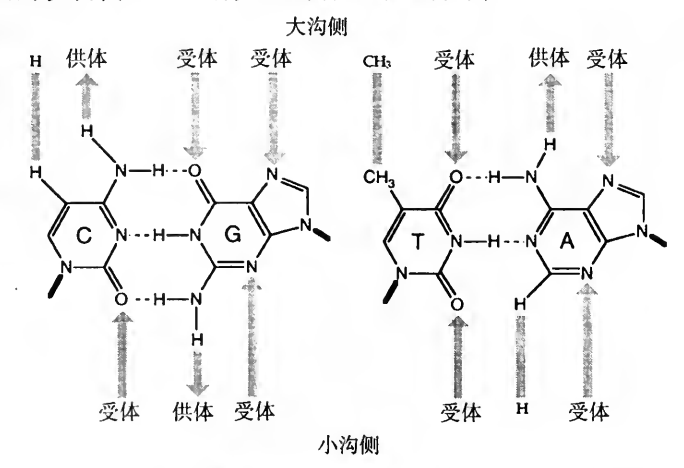


图 17.3 在大沟和小沟中 DNA 碱基配对的分子性质。大沟允许对所有四种可能的碱基配对加以区分,因为键的分布是不对称的。相反,在小沟中键的分布是对称的,所有很难区分 A:T 和 T:A 以及 G:C 和 C:G 碱基对。粗键表示了每个碱基与 DNA 骨架之间的连接。

解读蛋白中的氨基酸 在 DNA 结合蛋白中一些保守的 DNA 结合结构域的发现很自然使我们提出这样的问题:是否存在有一套特殊的规则指导着序列特异性的识别,例如,核苷酸识别的一个氨基酸密码。在大的蛋白家族(HTH, 锌指,基本结构域)中存在着不同的序列特点却有着结构上十分相似的结构域的例子。这样可以在三个水平上进行精细

的突变分析:①结构域交换;②部分交换(置换);③结构分析。

结构域交换实验包括在不同的蛋白中进行 DNA 结合结构域的交换。这种实验已经应用于在许多例子中,它可以了解这些 DNA 结合结构域对于普遍 DNA 的结合能力和序列特异性的结合是否是必需和足够重要的。分离的结构域常常可以单独结合 DNA,而杂交蛋白同样具有其原 DNA 结合蛋白的结合特点。

部分结构域交换是最精细的水平,包括单个残基的置换,可以鉴定出控制序列特异性的 DNA 结合结构域中的某些特定区域。置换特殊残基可以改变识别特异性,在比较那些在结合特异性上不同而具有高相关性的蛋白中,常常可以说明仅有一些必要残基涉及结合。根据其它蛋白而进行残基的系统性置换可以造成在结合特异性上发生变化,其变化后的特异性可同其模型蛋白相同。最近,系统置换已被*噬菌体展示*(phage display)(参阅)所替代:这种方法可以使 DNA 结合结构域随机地在一个或几个位置发生改变,这样变化丰富的突变体展现在噬菌体的表面,可根据识别特异性加以选择。这种技术已经成功地应用于锌指 DNA 结合蛋白的分析中,在下面将加以讨论。

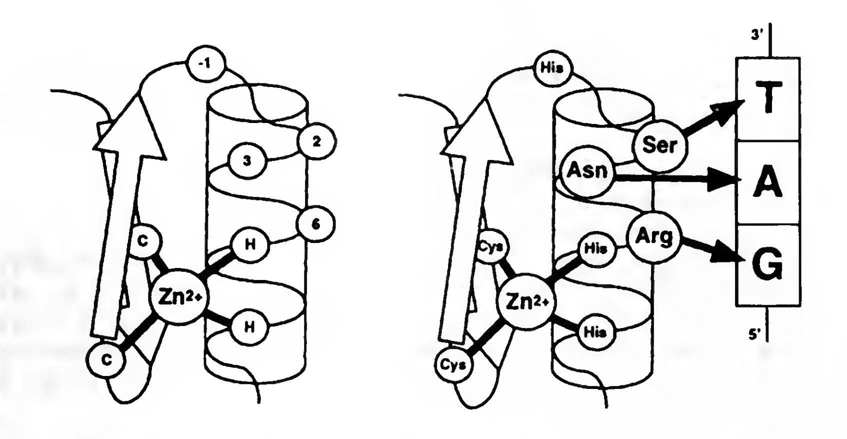
通过 X 射线和 NMR 对于蛋白-DNA 复合体的结构进行分析是在原子分辨水平上确定分子内键的空间排列的唯一方法。许多蛋白-DNA 结构已经被解析,这些研究在对于揭示 DNA 和蛋白在适应结合过程中的构象变化,以及水分子在蛋白-DNA 复合物中的作用中起着特殊作用。

假定的锌指蛋白 DNA 识别密码　　通过系统的结构域交换实验和蛋白-DNA 复合物的高分辨结构分析表明,没有一套普遍的密码适用于所有序列特异性蛋白和 DNA 的相互作用。这反映了识别结构的多样性,同时也反映了同样的识别结构,例如一个 α 螺旋,可以通过不同的方式呈递于 DNA。在 DNA 结合蛋白的一些家族中,相互作用是十分复杂和高度变化的,因为可由不同的方式产生同样的三级结构。同样的结构可介导序列特异性地又可介导同 DNA 普通地相互作用(例如,位点特异性重组 INT 和同其紧密相关的普遍 DNA 结合蛋白 HU),这便是一个很明显的例子。

然而,对于某些蛋白家族,同 DNA 的作用特点简单,因为总是同样的几个氨基酸侧链同碱基作用。这至少在一些 Cys_2His_2 锌指蛋白中是事实,同时其空间立体化学码已被确立(表 17.2)。DNA 和 Zif268 的锌指蛋白的相互作用已经被研明。识别螺旋中,三或四个特殊残基侧链在同碱基作用中起主要作用,而且大多数为简单的一一对应成键。锌指码让人联想到*遗传密码*(参阅),虽然有简并但并非含混不清。目前为止,这种密码仅基于很小一部分结构信息并且不完善。但是今后,随着越来越多的结构的信息被解析和吸收进来,将变得越来越复杂,并且在碱基识别中其他氨基酸侧链的作用也将被揭示。

表 17.2　与 Zif268 相关的锌指蛋白的部分 DNA 识别密码

	三联体密码中碱基的位置		
碱基	5′	中部	3′
A		3→Asn	-1→Gln + 2→Ala
C		3→Asn, leu, Thr, Val	
G	6→Arg;6→Ser, Thr + 2→Asp	3→His	-1→Arg + 2→Asp
T	6→Ser, Thr + 2→Asp	3→Ala, Ser, Val -1→Asn	-1→Gln + 2→Ser



碱基由特定的氨基酸识别,特别是在锌指组件(左图)的 α 螺旋识别位点的 - 1,2,3 和 6 位。例如果蝇的 Tramtrack 蛋白(右图)锌指 1 的相互作用。在很大情况下,碱基和氨基酸侧链的比例是 1:1。在其他情况中,两个氨基酸侧链相互支持来识别一个碱基。在这个表中一个碱基对应于螺旋上相应的位置的氨基酸是特异的,每个位点含有氨基酸 x、y 和 z。若两个残基相互支撑,则第一个残基后面加上"+"表示。碱基以三联体排列,不同的氨基酸用来识别三联体中不同位置上相同的氨基酸。这是由于识别螺旋的螺距是根据 - 1,3 和 6 位置上的残基,它们不完全处于螺旋的同一表面上,因此离 DNA 的距离也不同。与不同长度侧链接触的同一类氨基酸往往用来作为碱基的识别(如 Val, Leu, Ile)。象遗传密码一样,锌指的密码也是简并的,但含义明确。它也涉及碱基的三联体,但这是与锌指组件有关的较表面的性质,这个密码本身不象遗传密码那样涉及 3:1 翻译。

序列特异性的 RNA 识别　　相对来讲,序列特异性识别蛋白同 RNA 的相互作用的知识了解得较少,这主要是由于只有很少 RNA - 蛋白复合物的结构得以解析,这些结构,如 3 种氨基酰 - tRNA 合成酶,结合于 U1snRNA 发夹的 U1A 剪接体蛋白,和结合于 RNA 发夹结构的噬菌体 MS2 的衣壳蛋白,揭示了少数统一的原则。还有部分原因是因为 RNA 在细胞中的作用多样化,同时形成的结构也就多样。

氨基酰 - tRNA 合成酶是很独特的一类酶,它们被认为以吲哚异氰方式同其底物作用。但是这些结构的细微之处仅有很小的相似性,对此我们不必惊讶。我们根据是哪一类 tRNA 末端腺苷酸的羟基负载氨基酸,而将乙酰氨基酸 - tRNA 合成酶分为两类。第一类酶以 2′羟基作用,在这类中,大肠杆菌的 Glu - tRNA 合成酶的结构已被解析。而第二类酶以 3′羟基作用,其中啤酒酵母 Asp - tRNA 合成酶和耐热菌的 Ser - tRNA 合成酶的结构已经解析。每种酶以不同的方式结合 tRNA。Asp - tRNA 合成酶结合于受体茎部的大沟处,而 Glu - tRNA 合成酶结合于小沟的一侧(这两种酶有着不同的催化机制)。两种酶都同 tRNA 的反密码子环作用并识别反密码子,但他们的作用方式不同。Ser - tRNA 不同反密码子环作用,但是其螺旋卷曲结构可同 TΨC 环和可变环作用。这里包含了直接和非直接的解读机制。

17.6 研究蛋白核酸相互作用的技术

同蛋白作用的核苷酸序列的描述　　分析蛋白核酸相互作用常常从研究同蛋白结合的

DNA 或 RNA 序列开始。现在已发展了许多方法来鉴定和了解这些序列的特点,这些方法既可以将蛋白核酸复合物从裸露的核酸中分离出来,又可以在蛋白结合的保护中使核苷免受化学和酶的降解。这些技术的具体细节将在表 17.3 和图 17.4 中描述。

表 17.3 用于鉴定同蛋白结合的核酸序列的技术(图 17.4)

用于研究蛋白核酸相互作用的方法
技术:**凝胶阻滞分析**(gel retardation assay)(电泳迁移分析、迁移带分析、胶迁移分析)。 原理:DNA/RNA 蛋白复合物比裸露的 RNA 或 DNA 的电泳迁移得慢。 用途:鉴定和了解蛋白在 DNA 和 RNA 结合位点。 说明:同蛋白抽提物保温或未保温的标记的 DNA 和 RNA 片段用于凝胶电泳。成功的结合显示为迁移带由于胶阻滞作用,在经保温或未保温的样品之间出现不同的迁移带(蛋白核酸复合物有着相对慢的迁移率)。蛋白还可以进一步用抗体结合鉴定,由于阻滞变大,带将出现进一步迁移[或**超迁移**(supershift)]。结合位点的精确定位还可以通过加入过量的特殊碱基序列的**竞争性寡聚核苷酸**(competitor)加以研究。如果竞争结合位点是适合的,竞争性寡聚核苷酸将从标记的 DNA 或 RNA 处俘获那些蛋白,这样迁移带将消失;如果竞争位点不适合,将没有明显现象。这种方法可用于精确定位连续的结合位点。
技术:**DNA 酶 I 足迹法**(DNase I footprinting) 原理:结合有蛋白的 DNA 片段受到保护而不被酶降解。 用途:鉴定被特殊蛋白占据的准确的核苷酸位点。 说明:DNA 在相同或不同蛋白抽提物保温后,仅在一端被标记,同限制酶量的 DNase I 酶解产生一组标记条带。这些片段通过电泳产生一系列条带。蛋白同片段结合保护其不被酶切,在这些条带中产生间隔。同时对相同片段进行测序(参见 Maxam and Gilbert 测序)来确定被蛋白保护的正确的核苷酸序列。DNase I 足迹法可以同时鉴定同一 DNA 片段上的几个结合位点,并且竞争核苷酸也可以同胶阻滞中一样使用来了解其结合的特异性。
技术:**修饰保护**(modification protection)[例如硫酸二甲酯保护分析(Dimethylsulphate protection assay)] 原理:结合 DNA 或 RNA 的蛋白可以保护特殊的碱基不被化学修饰(参见修饰干涉 modification interference) 用途:鉴定同蛋白作用的特殊碱基 说明:硫酸二甲酯(dimethylsulfate,DMS)可以特定地甲基化鸟嘌呤碱基,作用后可以用哌啶在修饰位点处剪切。蛋白可以结合鸟嘌呤碱基而防止甲基化而不被剪切。DNA/RNA 同蛋白抽提物保温后,用 DMS 处理,同裸 DNA 的剪切产物相比较,被保护的鸟嘌呤便被鉴定出来。DMA 保护也可以用于细胞内来了解 DNA 与蛋白的相互作用(**体内足迹法**)而不仅仅用于体外研究。DMS 对于细胞是可透过的,主要可以直接甲基化胞内基因组的鸟嘌呤,接着将 DNA 分离,用哌啶处理,再用 PCR 加以扩增来鉴定被保护和未被保护的产物。如果 PCR 产物位于被保护的鸟嘌呤的两侧,如果有蛋白结合可得到额外的 PCR 产物。
技术:**修饰干涉**(modification interference) 原理:碱基被化学修饰后通常不会结合蛋白(参见甲基保护) 用途:了解同蛋白结合的特殊碱基 说明:在 **DNA 甲基干涉分析**(methylation interference assay)中,DNA 片段用限量的 DMA 处理,这样甲基基团随机引入,然后同蛋白抽提物保温,通过凝教阻滞电泳分离。如果一个特殊的鸟嘌呤碱基对于蛋白结合是必需的,修饰甲基可以影响结合并且所有包含修饰碱基的片段将在未受阻滞的裸 DNA 带中出现。将受阻滞和未受阻滞的条带分别割下,DNA 用哌啶处理。由于未被阻滞的条带中含有唯一的片段,其大小可以反映可结合蛋白的碱基的位置。使用同样的方法,而用不同的化学修饰剂,可以了解 DNA 中其他碱基的作用。对于 RNA,焦碳酸二乙酯(diethylpyrocarbonate,DEPC)可以羧基化嘌呤,并且肼可以剪切嘧啶。这样不仅可以阻止 RNA 同蛋白质相互作用,而且 RNA 可被苯胺剪切。

同核苷酸作用的蛋白质的纯化方法 用传统的蛋白纯化方法(如 HPLC)纯化核酸结合蛋白,如转录因子的结果往往较差,这是由于这些蛋白的丰度往往较低。然而,利用蛋白同特殊的 DNA 或 RNA 序列有结合的特点,可以得到少量十分纯的核酸结合蛋白。首先,用蛋白抽提物同总基因组 DNA 或非特异性的 RNA(如酵母 tRNA)混合去除非特异性结合的蛋白,然后,根据共有 DNA 结合序列或特殊的 RNA 序列设计的寡聚核苷酸用于提取感兴趣的蛋白。一种早期用于分离 RNA 结合蛋白的简单方法仅是将混合物通过

硝酸纤维素层析,因 RNA(包括其结合蛋白)可结合于纤维素上,而其他的蛋白被洗脱除去,这样 RNA 结合蛋白可用阻滞样品分析中得到。一个高效并应用广泛的分离 DNA 结合蛋白的方法是亲和层析,它是将寡聚核苷酸连接于层析柱中的固体基质上,再将抽提物上样,在这个过程中,结合蛋白将被吸附。使用生物素－链霉素系统(biotin-streptavidin system,参阅)的亲和层析过程同上。

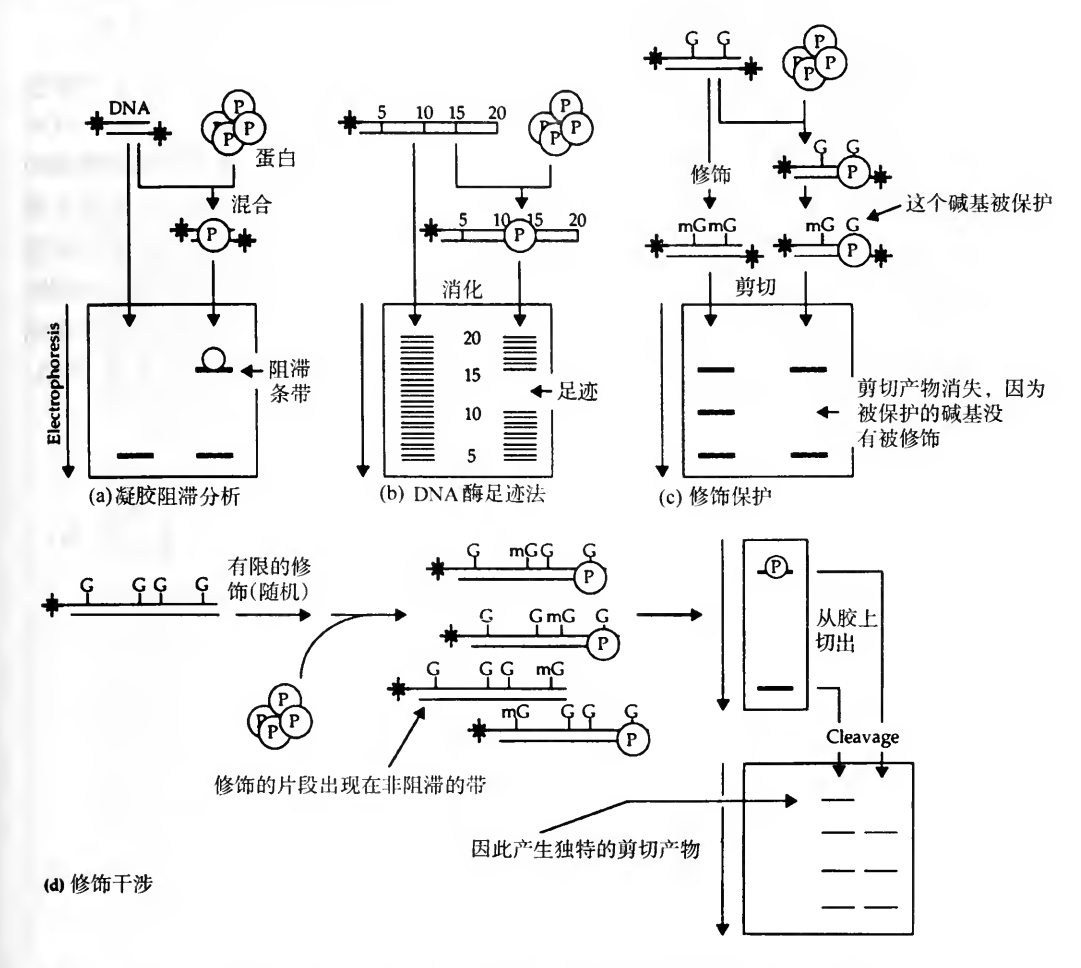


图 17.4 分析 DNA 同蛋白的相互作用。标记 DNA 或 RNA 同蛋白抽提物混合保温,用图示方法并且用电泳分离。具体方法可参看表 17.3。(a)**凝胶阻滞分析**:蛋白核酸结合的迁移带显示,带迁移率不同是由于蛋白核酸复合物被凝胶阻滞所致。(b)**DNA 酶 I 足迹法**:可通过在 DNA 的一端标记,再用限制酶量的 DNA 酶 I 处理成一系列的标记片段的方法,精确定位蛋白结合位点。蛋白结合片段受到保护而不被酶解,因而在梯带中产生间隔。(c)**DMS 保护**:DMS 可以特异性地甲基化鸟嘌呤碱基,使之可被哌啶剪切。结合蛋白可防止鸟嘌呤被甲基化,从而防止被剪切。结合特定鸟嘌呤的蛋白可由缺失的酶解片段所显示。(d)**DMS 干涉**:同 c 法,在蛋白保温前进行修饰,因此蛋白结合被阻止,裸露的 DNA 从 DNA 蛋白复合物中分离出来,从而分离得到包含鸟嘌呤的片段。这个片段可以和蛋白结合成非阻滞带。将各个片段分别用哌啶剪切,进一步电泳,通过产生的额外剪切产物显示出鸟苷带型。

用上述方法纯化的蛋白可通过 Edman 降解(参阅)进行微量测序,可用得到部分氨基酸序列,这样可以根据此设计简并的核苷酸探针或引物,通过筛 cDNA 文库分离对应于 DNA 结合蛋白的克隆。同样,还可以用含有核苷酸识别序列的寡聚核苷酸探针筛 cDNA 表达文库来获取基因(这个方法称为 DNA 结合蛋白的 **southwestern** 筛选和 RNA 结合蛋白的 **northwestern** 筛选,参见核苷酸杂交)。一旦 cDNA 分离得到,将其克隆于表达载体中,并在大肠杆菌中大量表达重组蛋白,这样利于更深入的功能研究(参见表达克隆)。

鉴定必需的核酸蛋白　　RNA–蛋白的相互作用反映了在细胞中的广泛功能,而核糖核蛋白常常是复杂的结构,包含几个 RNA 分子和许多蛋白分子(参见核糖体、剪接体、信号识别颗粒、编辑体、信息体)。通常通过确定他们的结构组成(例如某一成分同其余成分的作用)来研究核糖核蛋白复合物的结构和不同成分的功能重要性。一般可以用在文框 22.2 中所讨论的生物物理的方法来了解复合物的结构,还可以用分子交联,免疫共沉淀和亲和吸附的方法了解不同成分的相互作用,在功能研究中,可以采用在某个特定时间除去某一成分的方法了解其在复合物中的作用。RNA 可以 RNA 酶 A 去除,而特殊的 RNA 和其部分片段可先同寡聚 DNA 杂交,再用 RNase H 酶切,可特定地水解在 DNA、RNA 复合体中的 RNA 片段。

(袁　翀 译)

第18章
癌基因与癌

基本概念和定义

- **癌症**是多细胞生物的疾病，这些生物的细胞生长异常，失去调节，常伴有异常分化(**形成瘤，**neoplasia)。癌症往往是由于突变的积累而引起的(见下文)；它在高等脊椎动物中普遍存在，其他生物中因为生活期短，所以不发生癌症。动物中的癌症被广泛的研究，植物中也有肿瘤发生(参见冠瘿病，crown gall disease)。
- 癌症发生与自然选择类似，但只涉及个体中的体细胞而不是整个器官。正常情况下，体细胞有一发育过程，为了生物的生存和各种需要，需确保生长的控制。单个细胞失去生长控制时，要比周邻细胞生长快数倍，形成定域增生细胞群。在**肿瘤形成**(oncogenesis or tumorigenesis)过程中，发展成分隔的生长(**瘤**，neoplasm；或**原生肿瘤**，primary tumor)，可能会有未定性的分化细胞类型，并可进入血管等。肿瘤如果局限在一区域内是**良性的**，因为它通常可以通过外科手术切除。**恶性**肿瘤是**转移性的**——细胞破裂，播散到其他组织。在此阶段，癌症变得难以控制最终导致死亡。因为它会破坏和危害其他重要的器官。
- 癌症是由控制细胞生长的机制被破坏而引起的，或是由于蛋白功能水平上的病毒活性(偶尔也由细菌信号引起)，或是重要的生长调节基因的突变。有两类基因已被发现：**癌基因**(oncogene)促进细胞增生，癌症由它们不恰当的活性(产生功能突变)引起；另一类**抑癌基因**(tumor suppserssor gene，TSGs)抑制细胞生长，癌症可由于抑癌基因活性的下降(功能缺失突变)引起。原则上，有很多控制细胞增生和分化的信号传导系统和调节机制的破坏都会引发癌症(参见信号传导，细胞周期，发育：分子机制)。其他被发现与癌症有关的基因往往是在DNA修复或细胞对DNA损伤作出反应的基因。虽然这些基因的突变本身不会引起癌症，但它们增加了突变发生率，使癌基因和TSG突变更易发生。
- 癌症发生有几个阶段，经常从一个温和的生长紊乱开始(**发育异常**，dysplasia)，可以逐渐变得严重(**肿瘤发展**，tumor progression)。同样，原代培养的细胞开始也象体内细胞一样(它们表现出血清依赖，接触抑制、有限的寿命等)。它们可以通过一个多阶段的**生长转化**过程(growth transformation)，失去这些抑制特性，而转变成癌细胞。在这种过程的早期，细胞变得永生化，但其他方面仍正常。然而以后，细胞失去了血清依赖性等，逐渐变成致病性。这种逐渐严重的过程反映了控制细胞生长和分化的单一突变往往不足以引起癌症。癌症是突变积累引起的，它们作

用是逐渐扩大和以逐步方式(**多击假说**,multiple hit hypothesis)进行。这也反映了存在多种细胞增生调控机制,是动物生存的安全方面的需要,如人有大于 10^{13} 个细胞和很长的生存时间。每代每个基因发生突变的概率是 10^{6} 分之一,一个突变就引起癌症会使每个个体中数以千记的细胞受影响,物种不可生能生存。在任何一个细胞中癌的发生需 4—6 个独立的基因突变,使危险性大大下降。

18.1 癌基因

癌基因活性的基础　**癌基因**是有潜在的促发肿瘤发生活性的基因。它们最早是在分析**肿瘤病毒**(tumor viruses)中发现。肿瘤病毒的感染可以诱发肿瘤的生长。DNA 和 RNA 病毒通过不同的机制诱发肿瘤生长。DNA 病毒编码的产物通过在蛋白水平上影响宿主细胞的生长调节,这些产物在宿主的基因组中没有对应物(见下文)。RNA 病毒通过两种途径刺激细胞增生,或是携带一个高活性形式的宿主基因的变体,其本来的正常功能是促进细胞增生的,或是用病毒中强调控元件整合在邻近位置,激活内源性基因。在以上任一情况下,对应宿主基因被鉴定,这些基因突变时也有诱发肿瘤的能力(没有病毒的情况下)。

病毒癌基因和原癌基因　有些 RNA 肿瘤病毒可以在感染它们的宿主后立即引发肿瘤,被称为急性转化反转录病毒(acute tranforming retroviruses)。肿瘤发生是由于病毒基因组中存在的信息所引起,这种信息常由单个**病毒癌基因**(称为 v-*onc*)组成。癌基因常以牺牲病毒基因组信息为代价,急性转化反转录病毒是有缺陷的,需要**辅助病毒**(helper virus)来提供它失去的功能。第一个被发现有癌基因的反转录病毒是**劳斯肉瘤病毒**(Rous Sarcoma virus),它是个例外:除了有整个病毒基因组外,还有 v-*src* 基因。

每个病毒癌基因都有一个细胞内对应体,有类似或相同结构的**细胞癌基因**或称为**原癌基因**(以(c-*onc*)表示)。原癌基因是细胞内基因,功能是传导信号,调节细胞生长和分化。病毒癌基因是细胞序列被转导、修饰的复制品,并有不恰当的活性。原癌基因根据其产物的细胞功能可以分成多种类型(见表 18.1)。促进细胞生长的多种信号途径另行讨论(见信号传导),而癌基因的功能与这些途径中的成分相对应。原则上,任何信号蛋白不恰当的活性都有潜在的肿瘤发生能力。然而特别有趣的是,大多数是这些途径中的下游成分,直接控制基因的转录因子负责调控生长,它们汇集到细胞表面成为复杂信号,这些因子有的单独起作用(Ets、Myb),有的和不同癌蛋白(Fos 和 Jun)或其他蛋白(Myc 和泛蛋白 Max)形成复合物起作用。很多有诱发肿瘤能力的转录因子的表达量很低,由蛋白水平的磷酸化调节。突变引发过量表达或失去调节才具有诱发肿瘤活性。

因为反转录病毒只能在分裂期细胞中复制,而癌基因使它们诱发宿主细胞生长,所以这些病毒更多的是转导癌基因。一些癌基因最初以病毒癌基因形式被发现,对应的原癌基因的功能列在表 18.2 中。

表 18.1　癌蛋白的主要功能和编码它们的原癌基因举例

癌蛋白的功能	举　例
分泌蛋白(生长因子)	c-*sis*, *wnt*1
跨膜受体	c-*kit*, c-*erbB*
接头蛋白	*crk*, *vav*
GTP 结合蛋白	c-*Hras*
细胞内激酶	c-*abl*, c-*src*, c-*raf*
转录因子	c-*jun*, c-*fos*, c-*myc*
转录延伸因子	*ell*
RNA 结合蛋白	*ews*
细胞周期调节因子	*cycD*1, *cdc*25*A*

最先在病毒中发现的癌基因定为病毒癌基因(如 v-*myc*),定为细胞癌基因用前缀 c 表示(如 c-*myc*)

表 18.2　在急性转化病毒中发现的癌基因和它们在细胞内对应物的功能;大多数以病毒命名

癌基因	转导的反转录病毒	物种	原发肿瘤	产物[a]
abl	Ableson 白血病病毒	小鼠	淋巴瘤	酪氨酸激酶(cytosolic)
erbA	鸟红细胞瘤病毒	鸡	白血病,肉瘤	转录因子(甲状腺素受体)
erbB				受体酪氨酸激酶(EGF 受体)
fms	猫肉瘤病毒	猫	肉瘤	受体酪氨酸激酶(M-CSF 受体)
fos	FBJ 小鼠骨瘤病毒	小鼠	软骨肉瘤	转录因子
fps	Fujinami 肉瘤病毒	鸡	肉瘤	酪氨酸激酶(cytosolic)
jun	鸟肉瘤病毒	鸡	肉瘤	转录因子
kit	猫肉瘤病毒	猫	肉瘤	受体酪氨酸激酶(青灰因子受体)
mos	英洛尼鼠类肉瘤病毒	小鼠	肉瘤	丝氨酸/苏氨酸激酶
myb	鸟中幼粒瘤	鸡	白血病	转录因子
myc	鸟中幼粒细胞瘤病毒	鸡	中幼粒细胞瘤,肉瘤	转录因子
raf	小鼠肉瘤病毒	小鼠,鸡	肉瘤	丝氨酸/苏氨酸激酶
H-*ras*	小鼠肉瘤病毒(Harvey 株)	大鼠	肉瘤,红白血病	Ras GTP 酶
K-*ras*	小鼠肉瘤病毒(Kirsten 株)			
rel	网状内皮细胞病毒	火鸡	白血病	转录因子
sis	猴肉瘤病毒	猴	肉瘤	生长因子(PDGF β 链)
sis	猫肉瘤病毒	猫		
src	劳氏肉瘤病毒(RSV)	鸡	肉瘤	膜相关酪氨酸激酶

大部分病毒都携带了单个的癌基因,有些携带得更多。如 AEV 带有 2 个,v-*erbA* 和 v-*erbB*,它们是不相关的。很多癌基因是某种病毒所特有的,而有两种不相关的病毒都携有 v-*sis*。有的病毒品系带有相同癌基因的不同类型,如 MuSV 带有 v-*ras* 的不同类型。

缩略语:EGF,表皮生长因子(epidermal growth factor); M-CSF,巨噬细胞集落刺激因子(macrophage colony stimulating factor);PDGF,血小板衍生生长因子(platelet-derived growth factor)。

[a] 有关这些产物在细胞生长中的作用,可参见第 28 章信号转导。

突变激活癌基因　尽管很多癌基因是在研究急性转化反转录病毒中被发现，但大多数癌症的发生并没有病毒感染。另一种鉴定癌基因的方法是直接分析肿瘤细胞的 DNA，可以从结构上寻找正常细胞和肿瘤细胞之间 DNA 的区别，也可从功能上，将肿瘤细胞的 DNA 转染培养细胞，分析转化生长的能力。小鼠 3T3 细胞是这种分析的理想材料，因为它们已经历了生长转化，积累了癌症发生所需的几个突变。**3T3 分析**（3T3 assay）鉴别了很多最早在研究反转录病毒中发现的癌基因和其他转化病毒中未报导的癌基因（如 *wnt-1*，*trk*，*ret*，*lck* 和 *bcl-2*）。

在比较对应的 v-*onc* 和 c-*onc* 序列，以及从正常细胞和转化细胞中的 c-*onc* 的序列，发现原癌基因诱发肿瘤是由结构上的改变（质的改变引起不恰当的活性）或表达量的增加（产生的癌基因产物增多引起量的改变）引起的（图 18.1，表 18.3）。当正常癌蛋白被其结构所约束时（如受体需要它的配体激活），质的改变会引起非常严重的后果，而当正常癌蛋白活性受其丰度调节时（如转录因子），量的改变就更重要。最简单的质变是基因突变，无论是点突变还是平截突变，改变癌蛋白的结构，使其在正常量时有过高的活性。量的改变可能涉及调节元件突变，增加基因的转录活性或增加基因 mRNA 或蛋白的稳定性的突变。另外基因表达量的增加也可通过基因扩增或将原癌基因易位到强异源启动子或增强子的控制之下。在人类，癌基因的基因扩增经常与 EGF 受体基因相关，如 *RAS* 和 *MYC* 或 11 号染色体含有多个癌基因（包括 *BCL1*，*INT2* 和 *EST1*）的区域。扩增经常是 p53 蛋白功能缺失的（见下文）的二级反应，p53 蛋白功能缺失使复制在双链断裂处发生。至少在 50 种癌症中发现易位将癌基因置于强转录调控之下。易位也能产生了**融合癌基因**（fusion oncogenes）：产物可能功能正常，但过度表达[如在非洲淋巴瘤白基特淋巴瘤（Brukitt's lymphoma）的 Myc 中]或可能有不恰当的活性[如 ABL/BCR 融合蛋白在慢性髓细胞白血病中（chronic myeloid leukemia）所起的作用]。大多数融合癌基因被发现是编码改变底物专一性的融合激酶或融合转录因子。

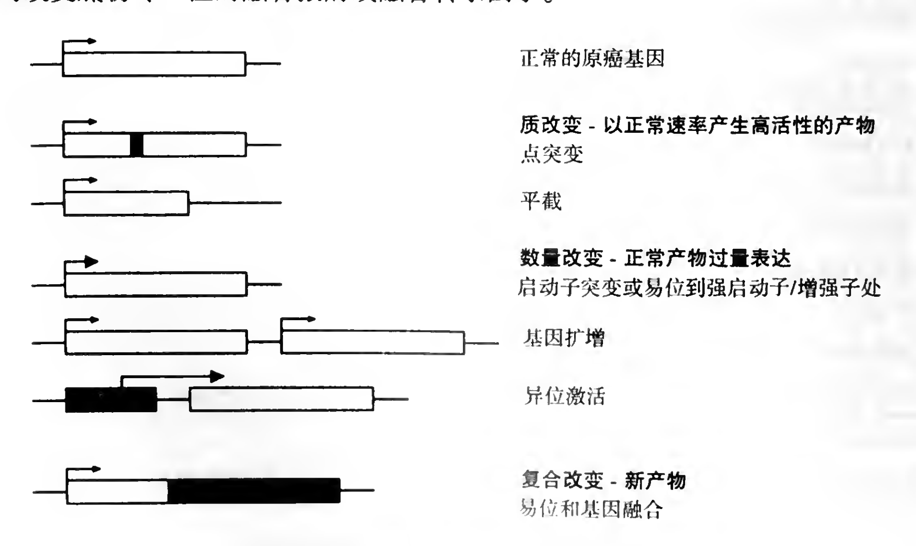


图 18.1　癌基因激活的机制。

表 18.3　癌基因激活的机制

改变	举例
质变	
点突变	c-*ras*： Ras是传导生长因子受体信号的GTP结合蛋白。受体激活通过募集鸟嘌呤交换因子，促进GTP与Ras的结合。Ras-GTP募集细胞内的Raf到细胞膜，使之激活并传导生长信号。Ras有GTP水解酶活性，使本身很快失活，所以它的信号传导活性是瞬时爆发的。*ras*基因的点突变，特别是影响12和61位氨基酸的突变，降低了GTP水解酶的活性，使其对生长因子的反应增大和细胞生长过量。
平截	v-*erbB*： c-*erbB*编码表皮生长因子受体，是跨膜酪氨酸激酶受体。病毒的癌基因v-*erB*在两端被截短，这样就同时缺乏N端配体结合结构域(使受体在没有配体存在下就形成二聚体)和C端的抑制结构域(使信号持续传导)。受体在没有配体存在下持续激活，使细胞增生不受调控。
量变	
扩增	c-*myc*： 特别是在生长受癌蛋白量控制的情况下，结构正常的原癌基因的简单扩增引起癌症。c-*myc*基因就是这种情况，c-*myc*编码的转录因子平时只有有限的量。这一基因在特定的癌中，包括乳腺癌，有数百倍的增加(参见DNA扩增)。
增加转录	v-*mos*，c-*myc*，c-*myb*： 在生长受某种癌蛋白有限的量的控制的情况下，癌症可由这种癌基因转录的增加而引起。这有多种方式。增加癌基因转录的调节突变在c-*myc*中有发生。另外病毒的转染使正常的原癌基因受一强病毒启动子的调节：转导的病毒v-*mos*与c-*mos*相同，由病毒的LTR启动子驱动而过度表达。原癌基因也可被整合在附近的反转录病毒反式激活。c-*myc*，c-*myb*和c-*raf*基因就是以这种形式被激活，在c-*myc*的基因座，反转录病毒的整合偶尔也会产生嵌合mRNA。
易位	
位置效应	*myc*： 在人类中，染色体易位t(8;14)(q24;q32)，将*myc*基因带到IgH的基因座，将其置于强免疫球蛋白增强子的控制下。这种移位被认为是白基特淋巴瘤的基础。MYC蛋白结构没有变化，但基因被上调(upregulate)了。与免疫球蛋白和T细胞受体基因座相关的易位常与癌症相关，因为V(D)J的重组体系提供了不正常重排的可能性，使基因座位于强增强子的控制之下[参见V(D)J重组]。
融合癌基因	*ABL*/*BCR*： *ABL*基因位于9号染色体，*BCR*基因在22号染色体上。交互易位产生一复合染色体(费城染色体，Philadelphia chromosome)，结果是*ABL*基因被剪接到不同长度的*BCR*基因的5′末端。融合产物缺乏ABL蛋白N端的激酶调节结构域，这样它的活性状态被BCR蛋白片段所稳定，结果是融合的酪氨酸激酶具有持续的活性。很多其他癌基因融合蛋白得到研究，其中大多数是转录调节因子。

不管通过何种途径，激活癌基因的突变是放大功能(gain of function)突变如促进生长活性的放大，并对野生型等位基因是显性。在反转录病毒转导的情况下被转导的癌基因必然比两份野生型的等位基因发挥更强的作用。

癌基因被慢性RNA肿瘤病毒激活　急性转化反转录病毒通过转导一高活性或过度表达的病毒癌基因到宿主的基因组中，使宿主细胞转化，而第二类RNA肿瘤病毒只能在感

染较长时间后启动肿瘤的发生。这些慢性反转录病毒不是缺陷的,也没有携带 v-*onc* 基因。癌基因的激活是因为反转录病毒整合到正常原癌基因的附近,使其被病毒中强启动子长末端重复序列所激活(参见反转录病毒)。这延迟的时期反映了反转录病毒的多轮复制和整合,直到病毒整合到某一位置,在此可以激活附近的原癌基因。不同的反转录病毒倾向于激活不同的原癌基因,如禽类白血病病毒(avian leukosis virus)倾向于激活 c-*myc*。这可能反映了每种病毒群反转录整合的区域专一性。

18.2 抑癌基因

抑癌基因活性的基础 肿瘤抑制基因(抗癌基因、隐性癌基因或癌症抑制基因),它的失活具有促进肿瘤生长的潜势。细胞融合实验预言 TSGs 的存在,一个生长转化的细胞可以通过与正常细胞融合得到恢复(即肿瘤生成表型得以纠正)。这样实验表明 TSGs 与一隐性失活的等位基因(如 *RB-1*)相一致,但很多 TSGs 突变后(如 *TP53* 和*WT* - *1*)(图 18.2)是显性失活(参阅)。TSGs 的正常功能是限制细胞生长,而且很多 TSG 的产物被发现是细胞周期(参阅)的调节因子或是生长相关基因的转录抑制子(表 18.4)。

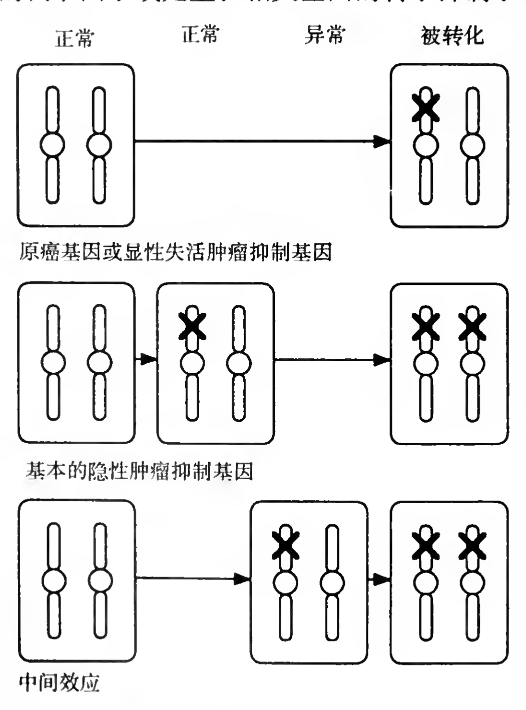


图 18.2 癌基因的突变和分类。

家族性癌症和努森(Knudson's)的两步打击假说 大多数由癌基因引发的癌症是由体细胞突变发生,而不是家族性的。这是因为癌基因是显性效应,在胚胎中失去生长控制的显性结构是致死的。家族性癌症按照孟德尔规律遗传,这也是存在隐性肿瘤抑制基因的证据。有意思的是,家族性癌症是不完全外显的常染色体显性遗传。

表 18.4　一些研究得较多的抑癌基因和它们的功能

TSG	蛋白的功能
RB1	成视网膜细胞瘤蛋白 RB1 在调控细胞周期中起关键作用。它与转录因子 E2F 家族和基本转录机器 RNA 聚合酶 I,III 的成分及重要的 UBF 结合并抑制其活性(参见 RNA 聚合酶 I)。当 RNA 聚合酶 I 和 III 合成 rRNA 和 tRNA 时,E2F 蛋白为进入 S 期而激活有关基因(如 c-*myc*,c-*myb* 和 *cdc2*)所需。RB1 蛋白控制细胞生长不仅通过阻滞细胞周期,还限制蛋白合成的速度。因此 RB1 受磷酸化调节。它在磷酸化状态是非活性形式,它在 S 期几小时前,被含细胞周期蛋白 D1 的细胞周期蛋白 - CDK 复合物(参见细胞周期)磷酸化。其他蛋白也与 RB1 作用,如细胞癌基因 *MDM2* 的产物。几种 DNA 肿瘤病毒癌基因的产物也抑制 RB1,不是通过隔离使 RB1 成为非活性形式的复合物,就是触发其降解(如 SV40 T 抗原)。
TP53	p53 蛋白是调节细胞增生和对信号(如 DNA 损伤)反应的调节因子,通过使细胞周期阻滞或诱导细胞凋亡来进行调节。细胞周期阻滞是通过转录激活编码 p21/p27 家族的 cyclin 依赖性激酶的抑制物以及抑制像 *MYC* 这类具有导致细胞增生活性的基因来进行的。p53 还有对 DNA 损伤识别和反应的广泛功能,因为它有 3′ - 5′外切酶活性,反映了它们具有修复 DNA 的功能。在对细胞肿瘤化的行为的应答中,p53 在特定细胞类型中可以诱导细胞凋亡,部分是通过对 *BAX* 和 *BCL2* 基因的调节(参见细胞调亡)。TP53 的点突变和缺失突变在所有人类肿瘤中有 60% 以上的发生率,并在 DNA 结合的结构域产生一连串突变。因为 p53 是四聚体,突变往往是显性失活的。稀少的家族性 Li-Fraumeni 综合征就与生殖细胞的 *TP53* 突变有关,特点是肿瘤谱系多样性。象 RB1 一样,p53 也可被 MDM2 癌蛋白抑制,并与多种 DNA 肿瘤病毒的癌蛋白相互作用。
NF1	NF 1 是 RAS-GTPase 的激活蛋白(GAP),它的功能是通过增加 RAS 内源的 GTP 水解酶活性,拮抗 RAS 信号。RAS 是鸟嘌呤结合蛋白,它与 GTP 结合时有活性,与 GDP 结合时失活。激活的 RAS 使 RAF 募集到细胞膜,在此处被磷酸化并能启动 MAP 激酶信号途径(见第 28 章信号传导)。*NF1* 缺失使细胞失去关闭 RAS 信号的能力,持续激活促有丝分裂的 MAP 激酶途径。当广谱的 GTP 水解酶激活蛋白 GAP^{p120}存在时,就会发生这种情况。生殖细胞 *NF1* 突变与神经纤维瘤和慢性髓细胞性白血病有关。
WT1	*WT 1* 基因通过不同的剪接,编码至少 4 个锌指蛋白。主要的剪接变异体是涉及生长调控的几个基因(包括 *BCL2*,*IGF-II* 和 *MYC*)的转录抑制因子。使 WT1 失去 DNA 结合能力的基因突变在肾的 Wilm 氏瘤和 Denys-Drash 综合征中可以发现。WT1 是以二聚体发挥作用,故 *WT1* 突变常是显性失活的。
P21 和 *P16*	p2 1 是 cyclin 依赖激酶的抑制因子(参阅),为 p53 对 DNA 损伤反应后激活。p16 是 CDK4 和 CDK6-D cyclin 复合物的特异性抑制因子。这些基因失活突变编码蛋白失去阻断细胞周期在 G_1 期的能力。D cyclins 负责促进 S 期的起始事件(参见第 2 章细胞周期)。
BRCA1 和 *BRCA2*	*BRCA1* 和 *BRCA2* 突变与常见的并发乳腺癌和卵巢癌家族性病例有关,其中一半病例只出现乳腺癌。这两个蛋白非常相似,都有一锌指结构,在细胞周期中受磷酸化调节。它们具体的功能还不得而知,但可能 BRCA1 在发育中发挥作用。

努森的被两步打击假说(Knudson's two-hit hypothesis)是对目前所存在矛盾的一种解释。第一个研究的家族性癌症是成视网膜细胞瘤,是一罕见的儿童视网膜肿瘤。成视网膜细胞瘤发生有家族性(60%)和散发的病例(40%)两类,而且家族性癌症是常染色体显性遗传性状。大多数散发的病例都是单侧的,而双侧的肿瘤多为家族性成视网膜细胞瘤。努森认为产生成视网膜细胞瘤细胞需要两个独立的突变(二击),在家族性成视网膜

细胞瘤中，胚系已有一次突变，所以体细胞只需一次突变就可满足要求。因为突变的发生率为 10^{-5}，而在视网膜中大约有 10^6 细胞，所以带有受精卵一个突变的个体很容易产生成视网膜细胞瘤细胞，肿瘤因此是强显性但外显率不完全。在散发病例中，一个细胞需要两次体细胞突变，这在每 10 000($10^{-5} \times 10^{-5} \times 10^6 = 10^{-4}$)人中发生一次。家族性成视网膜细胞瘤的显性遗传和外显率不完全反映了"第二次打击"体细胞突变的可能性，不是胚系丢失隐性功能等位基因而显示了显性效应。

两次打击假说为肿瘤细胞中经常性的**杂合性丢失**(loss of heterozygosity)所支持。如果以正常细胞(从血液中得到)为对照，成视网膜细胞瘤细胞作连锁标记经典分析，经常发现在对照中为杂合的标记在肿瘤细胞中是半合子。这是因为第二次打击突变往往是染色体缺失，将成视网膜细胞瘤基因座和相关的标记一起除去。尽管并不总是这样(第二次突变也可能是 *RB-1* 基因的点突变或是不影响侧翼标记的缺失突变)，杂合性的丢失可以用来鉴定和定位孟德尔规律的 TSGs 到一系列基因组上的标记如微卫星重复序列(参见第 8 章基因)。

DNA 肿瘤病毒　　大多数真核细胞的 DNA 病毒是裂解感染，导致宿主细胞死亡。偶尔特定的 DNA 病毒在不发生裂解感染而潜伏在细胞中。这些 **DNA 肿瘤病毒**[包括多瘤病毒家族(polymavirus，如 SV40)，人类乳头瘤病毒家族，乳头状瘤病毒(papillomavirus，HPV)家族和腺病毒家族(adenovirus)]能够转化宿主细胞。这是通过病毒癌蛋白的合成，作用和抑制宿主 TSG 蛋白实现的。肿瘤抑制基因的产物首先是通过它们与这些病毒蛋白的作用分离得到的，这些病毒蛋白与 RNA 病毒中的不同，它们在宿主基因组中没有对应物。多瘤病毒家族产生两种**肿瘤抗原**(tumor antigens)，T 和 t，它们是肿瘤产生所需的。已知 T 抗原与 RB-1 和 p53 作用并使之失活，使细胞周期从抑制状态中释放出。同样，腺病毒合成两种蛋白，E1A 和 E1B，前者与 RB-1 作用，后者与 p53 作用。HPV 病毒的 E6 蛋白也可抑制 p53 的功能，而 E7 抑制 RB-1。这样看来，P53 和 RB-1 在 DNA 肿瘤病毒的感染方式中起决定性作用，加上其他研究的结果，认为它们是细胞周期的中心调节因子。

(陈　淳 译)

第 19 章
细胞器基因组

基本概念和定义

- 在有性繁殖的真核生物中，父系和母系对子代核基因组的贡献都是均等的，产生典型孟德尔遗传模式(参见亲本印记和性连锁遗传)。所以**核遗传**(nuclear inhertance)被称为**双亲融合**(biparental)。
- 细胞核含有染色体，在有的真核生物中还有质粒。然而细胞中除了细胞核还有其他细胞器有自身的基因组(最主要的例子是线粒体和叶绿体)，同时细胞质中还有可能含有寄生虫，病毒或内共生体等有机体。
- 细胞质内遗传信息与细胞核内基因的遗传方式是不同的。**细胞质遗传**(cytoplamic inheritance，**核外遗传**或**基因组外遗传**)通常是**单亲遗传**(uniparental)，即仅由单亲提供到合子中。在动物中，细胞质遗传与**母系遗传**(maternal inheritance)是同义的，因为只有母系配子对合子提供细胞质。在植物中，母系遗传占主导地位，但某些物种是**父系遗传**(patennal inheritance)，还有的双亲都有细胞质传递的能力(不是同时起作用)。在植物和低等的真核生物中，双亲的配子都能为合子提供等量的细胞质，但是双亲之一细胞质的基因经常被选择性的破坏或失活，所以遗传方式还是功能性单亲遗传。后代的基因型只决定于双亲之一，与另一个的基因型无关(图 19.1)。
- 细胞器基因组在结构和功能上和某些真细菌的基因组的性质相似。因此一般认为它们是由寄居在早期真核细胞中的内共生生物演化而来的。细胞器基因组和核基因组在功能上的高度整合是细胞器功能所必须的，这也暗示着在细胞器进化的过程中，大量细胞器的基因都被转移到核基因组中去。

19.1 细胞器遗传学

母系遗传 母系遗传是指只通过母系进行基因以及它们所控制的性状的传递。这反映了一个事实:在很多高等的真核生物中，母亲提供受精卵全部的细胞质而父亲只提供父系的核物质。母系遗传的结果在图 19.1 中描述。两个对照种系的杂交实验中，其中一个等位基因是全显性的，正反交的结果对于核基因是相同的，而对母系遗传的基因是不同的，母系遗传取决于母系的基因型。母体遗传不能与母体影响(参阅)相混淆，后者只反映在早期发育中母体的控制，与核基因遗传相关，表现出正常的孟德尔分离比(见图 6.1)。

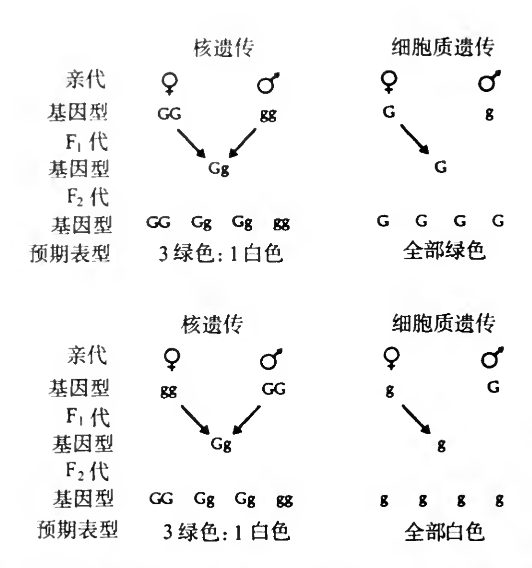


图 19.1 核(孟德尔)遗传和母系遗传。植物中颜色的变异往往是控制叶绿体功能的基因突变所造成的。其中一些突变涉及叶绿体携带的基因并表现为母系遗传。另外一些存在于核内表现为孟德尔遗传。如绿色(正常色素)对白色(失去颜色,无效突变)是显性的,就可以观察到不同的传递模式。孟德尔遗传中父母的遗传物质贡献相同,正反交的结果是同样的[植物并不具有与哺乳动物相似的异形性染色体,因而没有性相关的遗传(参阅)][1)]。

细胞器突变 影响细胞器基因组的突变可导致光合作用和氧化磷酸化的缺陷。叶绿体突变株经常有明显的形态表型(如斑叶),这在母系遗传的研究中已经得到了应用。真核微生物线粒体的突变被深入地研究,这种突变产生了生长较缓慢的表型[在酵母中称之为"小菌落"(peptite),在粗糙脉胞菌(*Neurospora crassa*)称之为"生长缓慢"(poky)]。在酵母中,细胞质遗传是双亲性的(但是在其他的真菌中不是这样)。这就使得对于线粒体缺陷株的研究是在异质状态下进行的(即多个不同基因型的细胞器共存于同一细胞质中)。在表 19.1 中根据酵母不同的"小菌落"突变体的遗传方式和它们对于野生型线粒体的显性进行了总结。表 19.2 定义了细胞器生物学中所用的一些术语。

人的线粒体突变表型是复杂的,因为它们导致组织特异性效应,这反映了线粒体 DNA 的产物与细胞类型特异核基因编码的产物相互作用。它们可能引起退行性的紊乱称为**线粒体疾病**(mitochondriopathies)。家族性的严重肌病,Kearns-Sayre 综合征,都和线粒体 DNA 的缺失有关,但是和酵母 *pet* 突变一样符合孟德尔遗传规律。这种疾病是由于控制线粒体 DNA 稳定性的核 DNA 的缺失造成的。另外一些疾病,如勒伯尔氏遗传性的视神经萎缩,与酵母的 *mit* 突变株类似,表现为母系遗传,反映了线粒体 DNA 本身的突变。在所有的人类线粒体疾病中,异质性的细胞同时含有突变型的和野生型的线粒体 DNA,突变型的细胞器对于正常的细胞器是显性的,这与酵母的 *rho* 的突变株类似。异质性与显性维持的机制还不清楚,有可能是突变线粒体 DNA 的自在复制,因为绝大多数突变型线粒体 DNA 保留了复制的双起始位点(见下文)。

1) 有些高等植物是具有异形性杂色体的,因而也存在伴性遗传,如女类菜属(*Melandrium album*)等,原著此处的提法有误。——译者注

表 19.1 酵母中线粒体的四种功能异常突变体

原有名称	突变型名称	分子基础
线粒体型小菌落 表现为细胞质遗传	*mit*	单个线粒体基因的点突变使其失去功能。
分离型小菌落 表现为正常的孟德尔遗传，缓慢生长表型	*pet*	影响线粒体功能的核基因发生突变。
抑制型小菌落 与野生型杂交后产生的全是小菌落	*rho*	失去大部分的 mtDNA，只剩下小的自在复制环，从而竞争超出野生型的线粒体。
中型小菌落 与野生型杂交后全是正常菌落	*rho*^{0}	失去全部的 mtDNA，线粒体的合成主要是核基因组控制，所以线粒体在失去所有的基因组后也能合成。

相似的分类也可见于叶绿体突变和动物的线粒体突变，反映了两个共同的主题。(1)细胞器和核基因组对于细胞器功能的双重控制。(2)在同一个细胞中，突变的细胞器对于野生型的细胞器是显性。

表 19.2 用于描述细胞器基因组行为的一些名词

名词	定义
细胞型	是由细胞质而不是核中的物质给予细胞的特性。不仅用来描述细胞质基因控制的性状，还可以用来描述在细胞质中合成并发挥功能的核基因所控制的性状(如参见 P 因子，杂种不育)。
同质的	由纯合细胞质而来－有同一基因型的细胞器的个体。
异质的	由杂合细胞质而来－具有不同基因型的细胞器的个体。
胞质杂合子	异质细胞。高等真核生物的细胞有同质的趋势，除非产生自发突变，胞质杂合子可以由细胞融合而产生。胞质杂合子在低等真核生物中是由双亲融合子的细胞质遗传而自然产生的。在这些物种中双亲的一方细胞质将被破坏，这一过程的失败就产生了胞质杂合子。这些异常的细胞叫做**双亲融合子合子** (biparental zygotes)。
细胞质分离	在有丝分裂中，异质性的细胞产生了同质性的细胞。这反映细胞器在子代细胞中的随机分配，一个随机的过程引起的失衡可以在下一轮的分裂中得到进一步的放大。一般几代以后，所有的子代细胞都是同质性的。相同的原理也可见于多拷贝质粒的分离(见第 20 章质粒)。
异核体，同核体，融合体	这些术语是与核而非细胞质基因有关的，但经常与细胞质遗传相混淆，因而在此说明。异核体是一个细胞含有一个以上的核且这些核有不同的基因型。同核体是多核的，但所有的核都是同一种基因型。融合体包含二倍体的核。一个合子可以被称为做融合体。

19.2 细胞器基因组

细胞器基因组的一般特征 利用多种技术来鉴定核 DNA 和线粒体 DNA 在结构和功能上的区别大大促进了对细胞器基因组分子水平的描绘。这些技术包括利用不同的抗生素选择性去抑制核基因或线粒体基因的表达，分离细胞器中转录和蛋白合成的分析，细胞器 DNA 和 RNA 的体外表达，连锁作图，限制性内切酶图谱，杂交分析和测序。细胞器基因组和核基因组在它们的结构组织，稳定性和基因表达和调节机制等方面表现出根本的差异。细胞器基因组的一些特性列于表 19.3。

表 19.3　细胞器基因组的一些特性

基因组经常是环状，并且是多拷贝的。

绝大多数细胞器基因都是有基因表达功能的（如 tRNA. rRNA. RNA 聚合酶）。还有的基因编码细胞器功能（如与光合作用和氧化磷酸化有关的蛋白），不过大多数蛋白是由核编码后运输进来的。

基因组经常以重组产生的序列变异体混合物的形式存在。

转录和翻译的调控是由原核生物类似的调控序列和反式作用因子完成的。

细胞器基因组的表达对于抗生素敏感，但对真核核基因的抑制因子并不敏感，这个现象对于核基因和细胞器基因分离研究很有用。

转录很复杂，包括多个转录起始位点和多顺反子信息。

细胞器基因基本上是以转录后加工的机制调控，包括 mRNA 稳定性的调节，RNA 加工，蛋白质合成和蛋白质降解。

基因表达具有特征性的复杂的 RNA 加工反应，包括切割，顺式-、反式-剪接，RNA 的编辑和降解。

细胞器经常使用统一的遗传密码的变体，反映更小 tRNA 种类的组成和更大的摆动相互作用（见第 11 章遗传密码）。

细胞器可以携带基因组外的质粒，通过介导基因重排质粒可提供宿主更多的表型。

质体基因组　原质体系（plastome）是从质体中分离得到的 DNA（质体是与叶绿体相关的多种植物细胞器之一）。对此 DNA 的分析发现，它们含有与**叶绿体 DNA**（chloroplast DNA）相似的 DNA（ctDNA）。它们的基因组为环状，大小在 110kb 到 150kb 之间，编码了叶绿体基因表达所需的各种蛋白和结构 RNA，包括各种密码子所需的 tRNAs，rRNA，一些核糖体蛋白和 RNA 聚合酶。此外叶绿体编码在光合作用中具有直接的作用的蛋白，包括光合系统 I 和 II 中的组成部分。叶绿体中所用的绝大部分多肽是由核基因组编码产生，再转运至叶绿体中。（参见蛋白质定位）。核/细胞质整合的程度非常高，在某些时候很多叶绿体编码的和核编码的蛋白相互缔合发生作用。例如，在一些植物中，核酮糖-1，5-二磷酸羧化酶-加氧酶（RuBisCO）二聚体的大亚基是由叶绿体编码，而小亚基由核编码。

在某些植物中叶绿体基因组基本上是单一的序列，但是在另一些植物中，是由两个单一序列及其中间插入反向重复序列组成，这些重复序列包含了 rRNA 基因等几种基因。在重复序列区经常发生重组产生插入型的异构体和多联体。

调控叶绿体基因表达的顺式作用元件在结构上与细菌的启动子类似，并可以在大肠杆菌中起作用。转录通常很复杂，产生不同大小的*多顺反子* mRNAs（参阅），在特定的加工位点加以剪切。基因的调控通常发生在 RNA 的稳定性和蛋白质合成的水平上，很多宿主编码的在叶绿体基因转录后调节过程中起作用的蛋白已经被分离出来。很多叶绿体的基因含有 II 型自剪接的内含子，这暗示还存在另外一些复杂的 RNA 加工的过程，如*反式剪接*和 RNA *编辑*（参阅）。蛋白质合成中的核糖体结合位点与大肠杆菌中的相类似，叶绿体使用的密码子是未修饰的普适的遗传密码。

线粒体 DNA：大小，基因组成和基因组结构　线粒体基因组是极其多样化的，其大小和结构组成在不同分类群中表现出极大的不同。虽然类似于叶绿体基因组，但线粒体基因组主要编码具有基因表达功能的基因（如 rRNA，tRNA 和 RNA 剪接酶）和一些有关线粒体功能的多肽。很多线粒体蛋白是在细胞质内合成后转运入线粒体。

酵母的**线粒体 DNA**（mitochondrial DNA，mtDNA）长度约 80kbp，是真菌 mtDNA 中

最长的。它包括了一些非编码富含 AT 的 DNA,它被富含 GC 的区域所分割,并含有复制起始位点。很多酵母 mtDNA 有很大的内含子,它们可能包含编码控制内含子剪接和转座蛋白的开放阅读框(参见归巢内含子)。在 mtDNA 中含有所有三种类型的自我剪接内含子(参阅)。tRNA 基因经常以功能簇的形式出现,而 rRNA 的基因是分散的。另外一些真菌的 mtDNA 比较小,大约 20kbp,含有较少的重复序列。

植物 mtDNA 在大小上表现出令人难以置信的多样性,从 100kbp 到 2.5Mbp,较大的基因组含有较高比例的 DNA 重复序列。在一些植物种属中,mtDNA 大小相同,但是很多植物中存在着复杂结构,包括不同大小的线状,环状分子和重组所产生的一些结构。完整的基因组称为**主环**(master circur);而小的衍生物叫做**亚基因组环**(subgenomic circles)。主环的重复序列为重组的位点——一般可以看见每一个基因组重组产物的全貌。异常重排 mtDNA(sublimons)的扩增,可能如酵母的 *rho* 突变株一样自在复制。

与此相对照的是动物 mtDNA。它往往比较小(小于 20kbp),结构上也很经济,不包含内含子,基因之间的空隙很小,只有单个控制复制和基因表达的非编码区。动物线粒体 DNA 的空间节约体现在多方面,包括全基因组转录(见下文),基因重叠和无终止密码子(它们是在 RNA 多腺苷酸化的转录后加工的过程中加上去的)。

线粒体基因表达 动物和真菌的 mtDNA 的转录是多顺反子的(参阅),真菌 mtDNA 中有多个启动子,但动物 mtDNA 中每一条链都只有一个启动子,最初表达为**全基因组**(pangenomic,整个基因组)转录本。植物中绝大多数的线粒体 mRNA 是单顺反子(rRNA 基因是例外),但它们是从多个起始位点开始转录的。所有线粒体都使用与原核生物相似的启动子序列,编码与原核生物酶相关的 RNA 聚合酶,包括使用与 σ 因子类似的辅助因子(参阅)。酵母和动物 mtDNA 的转录需要与 HMG 蛋白质家族有关的转录因子(参阅)。

线粒体中 RNA 的加工有好几种形式。植物中 rRNA 的转录本剪切后成熟,内含子被剪切(但是观察不到叶绿体中出现的反式剪切),一些基因可以进行 RNA 编辑。动物中,加工过程包括将多顺反子信使切割成单个的基因片段,tRNA 的二级结构基序为内切酶活性所作用,它的分散分布对于以上的过程可能起促进作用。动物线粒体的 mRNA 和 rRNA 都可以被多腺苷酸化。真菌的线粒体中 RNA 的加工与 3′端 12 个核苷酸信号有关,它们控制转录本的成熟。

植物,真菌和动物线粒体的翻译使用修饰的遗传密码,这些修饰是具类群特异性的,反映了较小的 tRNA 基因组成库,在某些情况下反映了特殊的 tRNA 结构。tRNA 种类的缺失可为摆动配对的修饰所补偿,所以可耐受更大的简并性(见第 11 章遗传密码)。植物的线粒体往往编码整套的 tRNA,虽然有些物种特定 tRNA 是被核编码翻译后转运入线粒体的。一些藻类和原生生物的线粒体有很少的 tRNA 基因,有的甚至没有,这种情况下,大部分 tRNA 是从核内转运的。翻译的起始包括原核生物类似的核糖体结合位点,起始氨基酸是 N-甲酰甲硫氨酸(见第 23 章蛋白质合成)。翻译是线粒体中蛋白质基因调控的重要层次。

线粒体 DNA 的复制 一些植物和真菌线粒体复制的起点已被阐明,但复制的机制还不了解:在有些植物中,控制复制的系统必须能够判别主环和亚基因组环(见第 26 章复制)。在动物中,由 RNA 聚合酶合成的大的 RNA 引物启动了前导链(**重链**或 **H-链**)的复制,引

物被核糖核蛋白内切酶 MRP 切割(参见核酶),然后 DNA 的合成就由 DNA **聚合酶** γ(参见)起始。引物的延伸和随从链(resident strand)的替代产生了 D-环,延伸经常在这一步被打断,这有可能与 D-环中的基序有关(**终止相关序列**,terminal associated sequyences)。这些位点的通读或被终止链的重新开始产生成功的复制。互补链(**轻链**或 **L-链**)的起始是在重链的对面,与一个特殊的线粒体引物酶有关。这种特殊的连续复制表明 mtDNA 在大多数时间内以单链形式存在,这可能是它突变率相对较高的原因之一(参阅)。

动核 DNA 原生生物含有单个的高度特化的线粒体称为**动核**(kinetoplast),与其他的线粒体相似,它有自己的基因组(**动核 DNA**,kinetoplast DNA,**kDNA**)。kDNA 基因组的结构有很鲜明的特点,整个的基因组[**大环**](maxicircles)在一个很复杂的网络中与很多小的衍生物[**小环**](minicircles)互连。大环和小环的复制都是使用滚环复制原理。kDNA 的原始转录本提供非常丰富的 RNA 编辑(参阅)的例子。一个极端的例子,就是 50% 以上的核苷酸被去除,引导 RNA(参阅)有利于这一过程的发生,它是小环合成的惟一产物。原生生物的动核(藻类线粒体也有)的特性还包括没有细胞器编码的 tRNA(见上)和小片段 rRNA 的合成,rRNA 散布于整个基因组。功能性 RNA 被认为是由分子内部碱基配对产生的。

细胞器质粒 很多植物的线粒体除了基因组外还携带质粒,它们的结构非常多样化(基因组是线性或双链环状 DNA,或单链/双链 RNA),它们很多与**细胞质雄性不育**(cytoplasmic male sterility,**cms**)有关,细胞质因子阻断雄性花粉的产生,这促进植物的异型杂交,因而总的来说对于植物是有利的。cms 表型可以为核基因的表达所抑制。例如 cmsT 的线粒体产生一个新质粒编码的蛋白,以一个我们目前还不了解的机制产生不育的表型。这种表型可以通过增加两个核基因座(*Rf1* 和 *Rf2*)的表达加以抑制,它们能够阻断此蛋白的合成。

真菌线粒体的质粒可能是隐蔽性的,也可能提供**长寿**(longevity)或**细胞死亡**(senescence)有关性状。后者是以细胞质的方式遗传,与为质粒整合所诱导的线粒体基因组的重排有关。例如脉孢菌的线性 kalilo 因子[1],虽然相同的表型可以由自主移动的内含子所引起,但它们中的一部分也可以质粒的形式存在。

内共生理论和混栖 DNA 细胞器有可能是由与原始的真核细胞形成内共生体原核生物演化而来的。这个模型的证据包括,细胞器与现有的细菌在基因组结构,基因表达的机制和对抗生素敏感性等方面的相似性,以及它们的基因和多肽序列的相似性。经历了一个进化时期,线粒体和叶绿体的部分基因丢失进入核内,从而对宿主依赖。**混栖 DNA**(promiscuous DNA)(细胞器起源的序列,可插入核或多种细胞器的基因组)的存在也支持了上述观点。混栖 DNA 的序列能够反映一些最近 DNA 转座的事件。虽然这种转座的机制还不清楚。但是基因从细胞质转移到核内的罕见事件的确发生。

细胞质遗传的其他形式 线粒体 DNA 和叶绿体 DNA 是最普通和了解最清楚的细胞质遗传形式。其他的细胞器虽然被确信也含有 DNA(如中心粒),但其编码潜势和功能

1) 脉孢菌中有一特殊品系,在一般培养中不能生长,不久死亡。在夏威夷发现,这是由于一线性质粒所致,夏威夷语称为死亡之门"kalilo。"——译者注

了解得还是很少。内源性的微生物以寄生和共生的方式与宿主共存时，也产生细胞质遗传的信息。我们已经讨论的一些事例如病毒，细菌甚至原核生物存在于大的真核细胞细胞质内，以母系遗传的方式，给予了宿主某些表型。在果蝇中发现了两个主要的例子，一个是病毒 sigma 提供二氧化碳的敏感性表型，另一个是寄生的原生生物的细胞产生特异性杀死雄蝇的毒素。这种性别调节器不传给或很少传给存活的雄性。

（张颂文 译）

第20章
质　　粒

基本概念和定义

- **质粒** (plasmid) 是在原核生物中常见的染色体外的自主复制子，它也存在于真核生物及其细胞器中。它们一般为共价闭合状态的超螺旋构型双链 DNA，长度为 1～300kpb。质粒可包含 1～100 个以上的基因。有些质粒可整合到宿主基因组中而被称为附加体。
- 质粒是寄生性的结构[*自在 DNA* (参阅)]，它可以通过接合、转化、转导或细胞融合等方式在群体间转移(见细菌中的基因转移)。质粒至少需要编码它在宿主中复制与维持所必需的产物。许多质粒还携带那些能促进细胞间自身转移的基因(参见*接合质粒*)。大部分细菌质粒还携有能提供对宿主有利但非必需的表型的基因(如抗生素抗性)。这些基因常常处于能在同一细胞中的不同质粒间跳跃的转座因子中。相反，真核生物细胞核中的质粒一般是**隐蔽性质粒**(cryptic plasmid，它们不提供表现型)并且其中有些似乎源于染色体 DNA。在质粒生物学中用到的术语列在表 20.1 中。
- 质粒和病毒的根本区别在于前者作为稳定的染色体外复制子存在，它们不编码病毒赖以形成感染性颗粒的外壳蛋白质。(因此质粒在细胞外只能以裸 DNA 形式存在)，一些病毒能以质粒形式存在，如 P1 原噬菌体在温和感染中以质粒形式存在，但反之不然。
- 质粒被广泛用作克隆载体 (见第 24 章重组 DNA)，但这一章的主要内容是介绍天然发生的质粒的行为。

表 20.1　在质粒生物学中使用的部分术语的定义

术语	定义
基本复制子	能复制同样大小完整质粒的质粒最小区域，典型的复制子含有：复制起点、**质粒维持序列** (plasmid maintenance sequences)，调节复制及质粒维持的元件(参阅)。
接合质粒与非接合质粒	即能或不能通过*接合*促使自身转移的质粒(参阅)。
拷贝数	每个细胞中质粒的平均数目，通常用质粒 DNA 数目或用编码基因产物的定量来量度。
***cop* 突变体**	携带能影响拷贝数突变的质粒。
隐蔽性质粒	没有明显表型的质粒
消除	自发或诱导性的质粒丢失，自发消除发生频率很低，但可通过插入剂和抗生素来诱导。

续表 20.1

术语	定义
排斥	一个常驻的质粒被另一个相容的质粒取代的罕见现象，这种现象常被解释为不相容性（参阅），但它也可能反映另一个质粒编码的限制性内切酶的活性等机理。
附加体	一种病毒或质粒既能在染色体外复制又能整合入宿主基因组中，如 F 质粒（见细菌中的基因转移）。
同质粒、异质粒	描述细胞携有一种类型质粒或两种不同类型质粒（参见*质粒分离*）。
不相容性	在对两种质粒都不进行选择的情况下，两种不同类型的质粒在几代后不能共存于同一细胞的现象，反映了共同复制或分配机制。
反转质粒	一种真核生物中线状质粒，带有较长的、完整的反向末端重复顺序。
杀伤系统	保证破坏被消除子细胞的维持系统（见表 20.4）。
维持系统	确保质粒在分裂的细胞群体中得以维持的系统（见表 20.4）
小质粒	较大质粒（如 F 质粒）中的基本复制子。
单向不相容性	1 型质粒被引入某一已有 2 型质粒存在的细胞群体中时，表现不相容性，但如将 2 型质粒引入 1 型质粒所在的细胞群体中，则不发生不相容现象，这反映了两种质粒相似但并不等同的复制机制。
分配，分配系统	分配是指在细胞分裂时，质粒分配给不同子细胞的现象，分配系统是一种保证平均分配的维持系统（见表 20.4）。
质粒	一种自主的染色体外复制子，它在正常生长条件下并非必需，也不是细胞基因组的一部分。
质粒复制起点	质粒 DNA 复制起始的座位。
质粒分离	在细胞分裂时，不同种类的质粒分离到不同的子细胞中（由于不相容）。
混栖质粒（广泛宿主质粒）	具有广泛的宿主范围的一种细菌质粒（通常包括革兰氏阳、阴性细菌）。
误切质粒（prime plasmid）	产生错误剪切并携带了部分宿主染色体的质粒附加体，**I 类误切质粒**（type I prime plasmids）质粒基因与宿主基因发生交换，**II 类误切基因**（type II prime plasmids）携带所有的质粒基因再加上宿主基因的一部分。（参见 F′质粒）
松弛型质粒	一种复制时不需要连续合成蛋白质的质粒，它的拷贝数在蛋白质合成受阻时仍继续增长，这是由于某种抑制性蛋白的去除。
严紧型质粒	一种复制时要求不断合成蛋白质的质粒，其拷贝数在蛋白质合成受阻时下降。

20.1 质粒的种类

根据表型分类　质粒可有许多种分类方法，最简单但也是用处最小的标准是它们赋予宿主细胞的表型（表 20.2）。在细菌质粒中，那些对复制、维持、转移不必要的基因经常位于可移动因子中，这些因子可以转座到其他质粒和宿主染色体中。由一种细菌质粒提供的表现型并不反映质粒分子本身的任何内在特性，同一质粒可提供多种不同的表型。**巨大质粒**（megaplasmids）携带着对多种抗生素抗性的基因因而对健康有重要意义。由于可移动因子的丰富性以及质粒重组的趋势，所以质粒本身的结构是流动的，因此设计出单纯以表型为基础质粒的命名系统非常困难（表 20.3）。许多真核质粒不具备表现型而无法用这一套命名系统命名，而另一些质粒，它们的表型并非来自其编码功能，而是来自于质

粒的行为(如染色体整合和重排)。

表 20.2 质粒提供的某些表型

表型	质粒及宿主种类举例
无(隐蔽性质粒)	啤酒酵母,2μ 质粒
抗生素抗性	肠细菌,R6k
抗生素合成	肠细菌,ColE1
抗原获得	耶尔森氏杆菌,Vwa 质粒
气泡形成	嗜盐菌,卫星 DNA 质粒
重金属耐受性	假单胞杆菌,FP2
携带宿主基因	啤酒酵母,3μ 质粒
长寿性	柄孢壳菌,PAL2-1
代谢利用	假单胞杆菌,OCT 质粒
限制修饰系统	混栖质粒,N3
衰老	粗糙脉孢菌(线粒体),*kalilo*
铁载体合成 (铁转运)	肠细菌(线粒体), 特定 Colv 质粒
不育性	玉米(线粒体) ,S1、S2 质粒
合成毒物	肠细菌,ENT P307
诱导肿瘤	根癌农杆菌,Ti 质粒
UV 保护(SOS 系统)	混栖质粒,R46

根据结构分类　虽然大部分质粒以双链闭环 DNA 形式存在,但质粒的定义并不排除其他结构。链霉菌属(*Streptomyces*)和梭状芽胞杆菌属(*Clostridium*)中已发现有单链环状的质粒存在,多种线状双链 DNA 质粒也已从一些细菌和真核生物中分离出来,如在包柔螺旋体属(*Borrelia hermsii*)细菌中的线性质粒编码不同的表面抗原,是回归热的发病原因。在含有线性质粒的真核生物中,哪些物质应被称为质粒,以及哪些应被称为染色体的界定并不清楚(即:哪些因子是染色体外物质,哪些是基因组的一部分)。从本质上说,在质粒与染色体间并无结构差别,认为细菌质粒在正常生长条件下是非必需的,这种常见说法是无意义的,除非所谓的正常生长条件的意义能明确定义。例如,酵母人工染色体及哺乳动物的 Y 染色体并不是必需的,但这两种因子通常不被定义为质粒。

生物中还存在 RNA 质粒。类病毒(参见)的特征为:它是不携带基因的单链环状 RNA 质粒。在酵母中发现的一些二重的、线状的、双链的 RNA 物质也符合这种质粒的定义。它们被称为**放毒者因子** (killer factor),因为它们使宿主表现出杀伤表现型(表 20.3)。它们编码一种包裹基因组的外壳蛋白,类似 RNA 病毒的衣壳蛋白,但它不能感染其他细胞而只能在细胞内转运,因此放毒者因子介于质粒和病毒之间,可以被划归于任何一种

表 20.3 基于表型的质粒分类与命名法

质粒类型	表型	例子
产细菌素	编码细菌素(能杀死或抑制其他细菌生长的蛋白质,如土壤杆菌素和大肠菌素),也编码免疫功能,以使宿主细胞不被破坏。酵母中的放毒者质粒(killer plasmids)就有类似功能。	Agk84,cloDF13 ,colE1 - K30 ,colv1-k94 以编码的细菌素命名(土壤杆菌素 84, 肠道菌素 DF13,大肠菌素 E1.V 及 la)
隐蔽型	无	2μ, *Mauriceville*
降解型	分解代谢	Lac,TOL,Cit,由底物命名(如乳糖,甲苯,柠檬酸)
F	致育性(接合转移)	F 质粒(见细菌中的基因转移)
R	抗性(如对抗生素,重金属)	R1,R46,RK6,多抗性质粒
重组(在体外由自然产生质粒部分构建)	通常为抗生素抗性——被用作主要选择标记(参见质粒载体)	pBR322,pML31,pBluescriptll,重组质粒通常冠以p开头的单词,后跟字母或数目以表示来源的实验室,商品化重组质粒给予俗名说明其用途。
致病力	能使宿主发生疾病,特别指那些编码直接致病功能的质粒(如毒素合成,肿瘤诱导)但也指非直接功能如抗生素抗性(可能通过使宿主抗药物的治疗而提高致病力)	Ent 质粒,Ti,Ri

这是一个人为的不基于任何质粒内在性质的分类系统,因而相关命名是不系统的,往往存在多种俗名,反映了很多质粒编码多种功能的事实。

(参见*亚病毒物质*)。注意,放毒者因子不同于放毒者质粒,后者是能使宿主呈杀伤表型的较常规的酵母 DNA 质粒。

根据内在性质分类　利用质粒的内在性质如其转移、复制和维持机制来分类是更好的质粒分类系统。

在细菌中,质粒可通过四种途径转移:细胞融合、转化、转导和接合(见细菌中的基因转移)。前三种过程对于质粒来说是被动的,而接合是主动的,许多质粒携带有可促进以这种方式在细胞间进行自身转移的基因,因此细菌质粒可以归成两个主要类别:**接合型**(conjugative)和**非接合型**(nonconjugative)。接合型质粒可根据特异的接合机制再细分下去(文框 20.1)。

在真核生物中,质粒转移机制在分型中用处不大。平行质粒转移往往只发生在细胞融合中(如有性生殖,或者真菌菌丝网的形成),或者偶尔通过机械转移(类病毒即通过此途径传播),偶尔细菌质粒也可转移给真核生物,像发生在细菌酵母之间的接合,以及根癌农杆菌属 Ti 质粒的特例中。

20.2 质粒复制和维持

质粒拷贝数　质粒复制和维持是所有质粒都具备的本质特性。因而可根据它们复制和分配策略及它们的拷贝数的调节方式来进行分类。

质粒复制是自主的,但可能与宿主基因组的复制相偶联,并可能受宿主及其所处的环

境影响。一般来说,质粒需要利用宿主编码的复制工具(如 DNA 聚合酶、RNA 聚合酶、DNA 引发酶),但编码复制起始及其调节所需的其他因子。复制起始于一个特殊的起点,这是质粒复制的限速步骤。成功启动的频率决定了每个细胞中质粒分子的特征性的平均数目,即**拷贝数**(copy number)。

在细菌中,较大的质粒通常有 1～3 个拷贝,它们的复制偶联于染色体的复制,被称为**单拷贝质粒** (single-copy plasmids),他们通常是接合型质粒。较小的质粒有较高的拷贝数 10～30,称为**多拷贝质粒**(multicopy plasmids)。他们的复制随机发生而且是自我调控的。这种质粒多为非接合型质粒。质粒复制的控制机制即保持适当的拷贝数的机制是通过衡量细胞中复制起点的相对浓度来实现的(见下文)。假如一个突变破坏了关键的调控元件的功能,则拷贝数将会有很大的改变。复制的强抑制可导致质粒丢失,而复制的非调节性复制(**失控复制**,runaway replication)可以提高拷贝数 10 倍。失控复制只有当 DNA 复制中另一成分,如一种酶或底物有限时才会受到阻遏。一个携带此种突变的基因被称为 ***cop*** **突变体**。

真核生物质粒的行为在啤酒酵母中进行了研究,带有 2μ 起点的质粒能保持高的拷贝数(<100),虽然复制只发生在 S 期(参阅)而且偶联于染色体复制。相反,类似于染色体的质粒(即拥有 ARS 起点和一个着丝粒)却维持 1～2 个拷贝数,着丝粒比质粒的复制起点更重要,所以人工合成带有着丝粒,ARS_s 和 2μ 复制起点的质粒其行为类似染色体(参见酵母克隆载体)。

质粒保持 如同自我转移和复制的调控,质粒也编码**维持功能** (maintenance functions),以确保在细胞分裂中两个子细胞均能获得质粒。将质粒分配到分裂成的两个子细胞的过程称为**分配**(partition),没有维持功能,在分配中就会发生自发消除(表20.1)。在正常生长条件下(即不存在对质粒所提供表型的选择)没有质粒的细胞比带有质粒的细胞具有更强的竞争力,因为它们不需将资源分流给质粒功能——因此,消除质粒的细胞在群体中快速增加,在数目上胜过带质粒的细胞。细菌中维持功能已被详细阐述并分成几个类别 (表 20.4)。维持功能的存在使质粒和病毒以及转座因子一样属于自在 DNA 种类 (参阅)。

质粒的不相容性 **不相容性** (incompatibility)是指两质粒不能共存于同一宿主株,除非供选择的表型必须由两质粒共同提供,否则不相容的质粒迅速地分离产生两种同质粒株。

不相容性发生在具有相似的复制和/或分离调控机制的质粒之间,质数拷贝数的调控方式主要是负调控。这样,在同一细胞中两个单拷贝质粒如有相同的复制机制,就会互相抑制复制直至细胞分裂为止,当它们分开后,抑制就解除了。混合多拷贝质粒也表现出这种的相互抑制复制的作用直至细胞分裂为止。在子细胞中,抑制作用被解除,而且由于多拷贝质粒的复制是随机的,每种类型的质粒都有相同的机会进行复制。得胜的质粒将获得较多的拷贝数,而且将在下一轮分裂中占优势,最终将导致同质粒细胞的增殖。因此多拷贝质粒将比单拷贝需要更多的时间才能分离。具有相同分配机制的质粒发生分离是由于二者竞争相同的"分配位点",这种位点,在细胞中可能只出现两次。质粒的不相容性提供根据复制和分配策略的有效的分类方法。相互不相容的质粒(具有相同策略)被放入一个**不相容组** (incompatibility group, Inc group),大约存在 30 类属于大肠杆菌及相关肠细

表 20.4 质粒维持功能

维持功能	描述
分配(*par*)系统	在低拷贝数的质粒中发现的系统,能精确均等地将质粒的拷贝在细胞分裂之前分离于细胞两极。分配系统需要反式作用因子和质粒中的顺式作用元件,可能还包括质粒与细胞膜相连的系带。细菌染色体使用一种相似的分配机理(见细胞周期)。高拷贝数质粒缺乏特定分配机理,依赖每个子细胞至少能获得一个拷贝的高概率。
杀伤(*kil*)系统(成瘾系统)	使质粒编码一种稳定的杀伤蛋白(killer protein,对宿主细胞致死的蛋白)和一个不稳定的"解毒"分子,这种"解毒"分子可作为杀伤蛋白本身的拮抗物或它合成过程的阻遏物。在被消除细胞中,杀伤蛋白比解毒因子持久,则细胞被杀死。
细胞分裂延迟	能在质粒拷贝数低的情况下延迟细胞分裂的系统。F 质粒能通过诱导 *SOS 反应*(参阅)来延迟细胞分裂。
重组系统	质粒编码了一种位点特异的*重组系统*(参阅),它能抑制导致多聚化的同源重组事件,并确保单体能被分配,大肠杆菌的 *xerC* 和 *xerD* 基因编码类似功能,以确保染色体的单体化,酵母 2μ 质粒也编码一个位点特异重组系统以提高其拷贝数。

注意:分配系统、杀伤系统和细胞分裂延迟系统均确保质粒分离的效果优于通过阻止消除细胞生成的随机分离。相反的,重组系统并不阻止消除细胞的生成,它们只能达到最佳的随机质粒分离。

菌属的不相容组(文框 20.1)。

质粒 DNA 复制机制 大部分细菌中的典型闭环双链 DNA 质粒的复制与染色体相似,它们的启动依赖于特殊的启动蛋白,即 **Rep 蛋白**(Rep proteins)与起点处的重复序列结合(见复制)。对于许多质粒(如 F、R1 质粒)Rep 蛋白促进复制起始位点的解旋,允许解旋酶的结合并促使单个或两个复制叉的形成。在另一些质粒中(如葡萄球菌中的 pT181),Rep 蛋白是一种*切口酶*(参阅),能启动*滚环复制*(参阅)。大肠杆菌相关质粒不需要 Rep 蛋白,因为宿主 RNA 聚合酶被用来产生一个通过复制起点引物起始引导链的转录。这种质粒处于松弛控制中(见下文)。

线状的细菌质粒类似于线状病毒基因组。例如:他们都有共价闭合末端(如 Borellia 质粒)或者末端蛋白和反向重复序列(如链霉菌属质粒),而且都可以利用与病毒相似的方法进行复制(见复制和病毒)。真核生物的线状质粒通常编码它们自己的聚合酶,据推测,这种酶可用于自主性复制。以下的讨论是关于质粒独特的复制调控策略。

多起点和重复区 很多较大的质粒都有多个复制起点(经常与不同的调控机制相联系),这样将这种质粒归入单个不相容组就很困难。在某些质粒中(如 F 质粒),某个特定的复制起始位点利用率较高,而其他起点的作用还不清楚。在另一些质粒中(如 R6K 质粒),不同起点有相同的使用机率。许多重组质粒载体已被精确地设计为既有真核复制起始位点又有原核复制起始位点,使它们能在两种类型细胞中维持(如*穿梭载体*)。

质粒复制起点往往以必需重复序列即**重复子**(iteron)为标志。象细菌染色体一样(参见*复制起点*),这些重复区就是质粒中的共同基序(如 RK2 质粒),也是一些能像质粒一样复制的噬菌体的共同基序(P1 和 λ1 细菌噬菌体)。质粒重复子是 Rep 蛋白结合位点,它既可单独作用也可以与宿主起始蛋白缔合起作用(如大肠杆菌中的 DnaA)。例如:在 RK6 质粒中,不管三个复制起始位点中哪一个被采用,γ 复制起点范围内的七个直接

串联重复序列对于复制都是必需的。

反义 RNA 控制质粒复制 带有一个 ColE1 类的复制起点的质粒（包括质粒克隆载体中的绝大部分），从一条单链 RNA 引物开始前导链 DNA 的合成，这条单链 RNA 引物由宿主 RNA 聚合酶合成（见引物，引发策略）。实际上复制起始位点处根据两条链转录出了两个转录本，RNAⅠ和 RNAⅡ。RNAⅡ是引物，而 RNAⅠ，与 RNAⅡ链部分互补，通过与 RNAⅡ形成失活的双螺旋而起阻遏作用。DNA 合成开始于复制起始位点，而 RNAⅡ转录始于复制起始位点上游 555bp 处，并且从那里开始延伸经过复制起始位点。此转录本必须在复制起始位点处被切割以产生功能的引物，这个过程由 RNaseH（参见核酸酶）催化，RNAⅡ所形成的三茎环二级结构促进了这一过程。

RNAⅠ和 RNAⅡ的相互作用在它们完全杂交（参阅）时达到顶点，二者杂交可以破坏 RNAⅡ结构并阻止它与 DNA 杂交。这反过来阻止了转录本的加工并引起复制阻断。在体外，只有当后者长度在 100～350 nt 时 RNAⅠ和 RNAⅡ能杂交，这意味着 RNAⅠ必须在 RNAⅡ转录本合成早期时与新生的 RNAⅡ转录本杂交以阻止复制。一个称为 Rom 的质粒编码的蛋白在 RNAⅡ长度超过 200 核苷酸时能增强 RNAⅠ：RNAⅡ杂交形成的速率，因而是 ColE1 复制的一个关键调节物。

R1 质粒的复制也由反义 RNA 调节。质粒编码的起始蛋白 RepA 是复制启动所必需的。但相邻基因的转录形成了一条反义 RNA（*copA*），它与 *repA* mRNA（称为 *copT*）的前导区结合，阻止 mRNA 的翻译（见反义 RNA，翻译控制）。*copA* 和 *copT* RNA 都可形成二级结构，突变分析表明两者间的相互作用可能包括这些二级结构，而不是全部双螺旋的形成[一个所谓的**相吻复合物**（kissing complex）]。

质粒中广泛存在着反义 RNA 调节物，这表明当蛋白质的调控不能完全满足需要时，RNA 为复制的调控提供了一种独特的策略。在许多不同质粒中存在着相同的机制：较短的反义 RNA 抑制因子与一必需的功能性 RNA 的 5′端杂交阻遏其活性。一个可能的解释是质粒拷贝数的维持要求一个不稳定的调节物，以便使质粒群落可以对标准拷贝数的偏离迅速作出反应。RNA 无疑是这种不稳定调节物的适宜的候选者。

松弛型和严紧型控制 质粒复制可能是在**松弛型**或**严紧型控制**（relexed or stringent control）之下。假如细胞的蛋白质合成受抑制，例如用氯霉素抑制，则一些质粒继续复制而且它们的拷贝数增长超过了一般营养状态的水平。这就是**松弛型质粒**（relaxed plasmids），是典型的小的多拷贝质粒。另一些质粒当新蛋白的从头合成被抑制时则伴随宿主染色体停止复制，称为**严紧型质粒**（stringent plasmids），是典型的大的单拷贝接合型质粒。大部分一般用途的质粒克隆载体都处于松弛型控制之下。

松弛型质粒在缺乏宿主蛋白合成时继续复制的能力反映了宿主复制启动所需的宿主蛋白的稳定性和质粒编码调节物的不稳定性。例如：ColE1 质粒处于松弛型控制下，因为启动因子 Rom 具有负调控作用——它通过促进反义 RNA 与前导引物链的配对来抑制启动。当 Rom 缺乏时，质粒复制解除抑制而且拷贝数增加（拷贝数很高的克隆载体没有功能性的 *rom* 基因）。相反地，严紧型质粒，如 F 质粒需要正调控作用的活性 Rep 蛋白促进复制起始。如缺乏 Rep，无论宿主 DNA 复制蛋白是否存在，质粒复制都会被阻抑。

文框 20.1　不相容组

与不相容性有关的问题　质粒根据它们相互之间的不相容性被归入**不相容组**(incompatibility groups, Inc groups),这种现象(不相容)表现为在一个包含*异质粒*(heteroplasmid)(参阅)的细胞群落中,随种群数目增长分离为两个*同质粒*(homoplasmid)(参见)的亚群,因为这些质粒有类同的复制或分配控制机制。但不相容的质粒可以表现出相容性,假如它们进行重组,或在拷贝数较高的情况下,数代后才会发生分离。相反地,相容的质粒能表现出不相容性,如果他们发生*排斥*(参阅),或假如这个细胞中有几个未被觉察的表现不相容性的隐蔽性质粒。总之,大质粒经常有多个不同的起点(**多复制子质粒**, multireplicon plasmids),并且不可能很清楚地归入某一不相容组。

不相容组命名法　在肠细菌中,不相容组用一个字母代表特定的质粒接合系统来命名,后加罗马数字和/或希腊字母顺序来表明不相容状态(反映分配与复制机理)。带有 F 型接合系统的质粒被分成两个不相容组, IncFⅠ和 IncFⅡ,它们有不同的复制控制策略,其他自然发生的接合质粒各有其独特的复制策略。来自非肠杆菌属的质粒也可以被归入不同的不相容组。在假单胞菌属中,不相容组包括:P1、P2、P3 组等,代表 15 个或更多不同复制系统,接合系统在这里不宜作为特征。在肠杆菌属和假单胞菌属的不相容组之间存在重叠,即存在广泛宿主范围的质粒,例如:P1、P3 和 P4 不相容组分别等同于肠杆菌属的 P、C、Q 不相容组。

以接合系统来划分的基础　在肠杆菌属中的接合系统,根据菌毛形态学进行划分(例如:硬的或有弹性的,粗的或细的),以及吸附在菌毛上的细菌噬菌体进行划分(见第 30 章病毒)。例如:F 型接合系统标志为粗的、有弹性的菌毛及其上附有类 M13 - DNA 和 RNA 噬菌体如 Qβ 和 R17。不相容组 IncFI 和 IncFⅡ代表两组 F 型接合系统质粒,它们具有不同的复制策略,第一种如 F 质粒本身,第二种如 R1 和 R100 质粒。

(殷向雷 译)

第21章
聚合酶链式反应(PCR)

基本概念和定义

- **聚合酶链式反应(PCR)**是体外扩增DNA序列的技术：即DNA复制不是像分子克隆技术中那样在细胞内进行，而是用纯化的酶进行。
- 基本的PCR是利用一对根据所扩增的目的片段所设计的寡核苷酸引物进行的。引物与对应的DNA链退火，DNA合成向中心区域进行。反应涉及三个不同温度的阶段：双链DNA的**变性**(90℃以上进行)，引物与单链模板的**退火**(50℃左右进行，最适温度根据引物序列而定)，**引物延伸**，合成新的DNA越过靶区域(70℃左右进行)。每组三个反应称为一个**PCR循环**(PCR cycle)，从理论上目标序列数量增加一倍(参见图21.1)。
- PCR反应在小体积(20～100μl)中进行，在单个PCR小管或多孔板中进行，用带有自动热循环的加热块来精确控制温度。反应体系中有模板DNA、引物、四种脱氧核苷酸三磷酸、热稳定的DNA聚合酶和反应缓冲液，其中最关键的成分Mg^{2+}离子。
- PCR的优点是比常规的克隆方法快速、灵敏，即使在DNA来源严重降解或从来源困难的DNA中分离(如已固定组织的情况下)都是非常有效的。但PCR也有出错倾向高，产物大小受到限制和必须预先有目标序列信息等缺点。

21.1 PCR反应的特异性

单一顺序的扩增　因为所有的DNA聚合酶都需要引物(参阅)来启动DNA的合成(参见复制一章)，通过设计针对目标序列的特异性引物，PCR可扩增特异性的序列。**PCR引物**是化学合成的*寡核苷酸*(参阅)，它们的设计是PCR的成功与否的关键。最有效的特异扩增反应的引物的长度是17～30碱基，因为这样的序列不太可能随机与高等真核生物单一顺序DNA(参阅)配对。引物经常有约50%的GC含量，避免同一种核苷酸串联和二级结构的出现，因为这样的结构会使引物成环和在错误的结合位点处稳定结合。除此以外，引物还要避免自身配对，因为这样会形成**引物二聚体**，造成产物的大幅度减少。

与通常的克隆方法相比，PCR有快速、灵敏的优点，即使在DNA来源严重降解或必须从困难的来源中分离DNA(如已固定组织的情况下)的情况下仍是非常有效的。但PCR也有出错，产物大小受到限制和必须预先有目标序列信息等缺点。

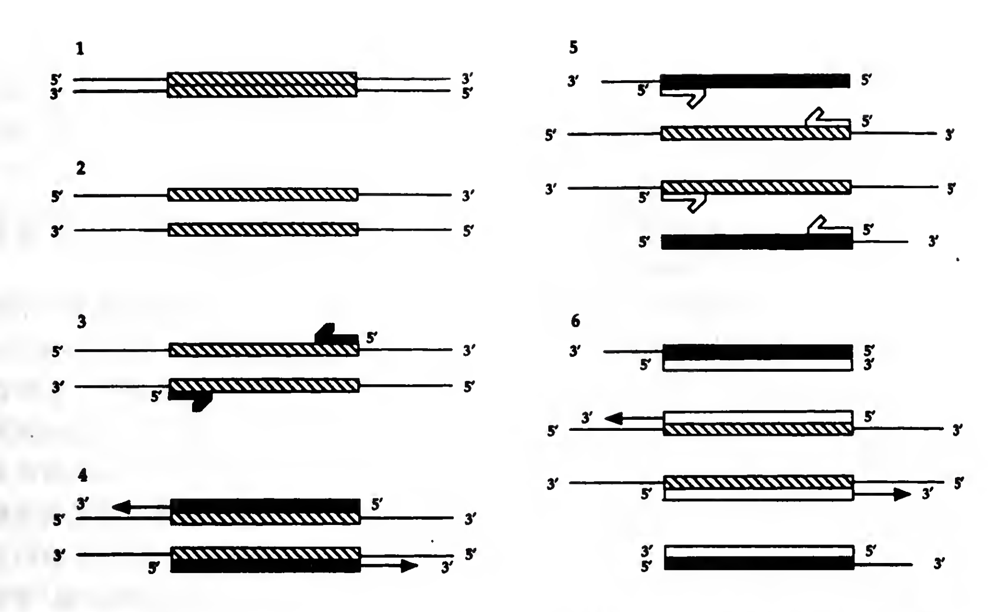


图 21.1 基本 PCR 反应。在第一个 PCR 循环中,原有的靶序列(1)以斜线表示被变性(2),引物退火(3)。引物向中心区延伸(4),使靶序列加倍,尽管第一个循环的产物的长度不具有专一性(以黑体表示)。在第二次循环中扩增的靶序列变性,引物退火(5),引物延伸(6)使靶序列量加倍。第二次循环的产物以白色框表示,其中一半大小没有专一性,另一半根据引物退火的位置,与扩增片段大小完全一致。在以下的循环中,这些特异性产物以指数倍积累,而长度不专一的产物以线性方式增加,在最终产物混合物中占很小的比例。

当引物在较低温度下退火时,可能引起特异性的降低,因为错配可被忍受并在引物延伸后被稳定下来。当反应成分在室温下混合时容易发生这种情况,但可以通过**热启动 PCR**(hot-start PCR)得到解决,热启动是当反应达到退火温度时才加入反应关键成分 DNA 聚合酶。**嵌套式引物**(nested primers)可以用来增加 PCR 的特异性,一次 PCR 扩增得到的产物接着用第二套引物进行第二次扩增。第二套引物和相同的靶位点结合但是在原来引物的内侧,所以在第二次反应中,只有特异性产物才能被扩增,称为**嵌套式 PCR**(nested PCR)。

单一顺序扩增的应用　　单一顺序扩增有很多方面应用。作为一种诊断技术可以用来判断在复杂样品中是否存在某一特定序列(如确定转基因或质粒插入的存在);检测多态性(如小卫星 DNA 的变异;参见 DNA 分型)和检测未知突变。特异性的序列可以作为**序列标记位点**(sequence tagged sites)(参阅),在基因组作图和定位克隆中起作用(见第 12 章基因组和作图,第 24 章重组 DNA)。引物无法结合也可用于诊断中。**等位基因特异性 PCR**(allele-specific PCR)可以用于检测疾病相关的点突变,引物根据突变位置设计,只能与特别的等位基因退火。这种检测方法比杂交分析(参阅)更灵敏和方便,因为它可以检测很小的序列差别,而这种差别在 Southern 杂交中不能检测到。此外,PCR 只需很少的材料,所以它在临床样品有限时非常有用,如绒毛膜取样(chorionic villus sampling)。

PCR 还可作为从复合物中扩增特异性序列而无需分子克隆的制备技术。这方面 PCR 的主要缺点是产物的大小限制。PCR 扩增较少被用于基因组克隆中,而用来产生部

分的 cDNA 片段，即表达序列标记(expressed sequence tag, EST)(参阅)。同样，这需要对目标序列有一定的了解，所以尽管 PCR 方法快速经济，但它不可能完全取代文库为基础的筛选方法，文库筛选提供了直接得到全长克隆的途径并可采用不同的筛选策略如免疫和配对筛选途径(参见表达克隆)。

相关顺序的扩增 PCR 的应用目的并不都是产生特异性的产物。PCR 也可用于鉴定序列相关的家族或根据一种物种的序列信息扩增另一物种的相关序列。

在这样的情况下，可以围绕着在高度保守的结构域来设计特异引物，使内部序列不同的产物得以鉴别。另一种方法是应用**简并寡核苷酸引物**(degenerate oligo-nucleotide primers, DOP-PCR)，简并引物是在某些位置上带有不同核苷酸的引物混合物。这种称之为**同源性筛选**(homology screening)的方法是根据保守结构域设计引物，但不与该家族中已知序列混同。这种方法已经成功地扩大了一些基因家族，包括 POU 结构域的转录因子(参阅)和细胞周期蛋白及细胞周期蛋白依赖激酶(参阅)。当只知道蛋白质的氨基酸序列简并引物还用于扩增对应于已知蛋白质的特异序列。在这种情况下，简并引物的应用反映了遗传密码的简并性，与简并寡聚核苷酸探针从 cDNA 文库中筛选克隆类似(参见第 24 章重组 DNA)。

非相关序列的扩增 PCR 的另一个主要应用是微生物的鉴别，如组织感染和水污染的检测，以便采用正确的处理和清除的方法。而根据形态学、代谢和行为表型的传统方法费力又只能用于少数种类。PCR 作为一种简单的诊断方法使用，可根据微生物的核苷酸序列获得明确鉴别。另一个优点是用多套引物可以在一个反应体系中同时检测许多不同种类的微生物(**多重 PCR**，multiplex PCR)，每一种都有特异性大小的诊断产物。

任意序列的扩增 在特定的情况下，PCR 的目的不是产生一个或多个特异性或相关的产物，而是为了收集任意序列的集合用于进一步分析。这种策略主要用于 cDNA 克隆和基因作图两方面。**任意引物**(arbitrary primers)进行扩增即用能和很多位置退火的短引物(9－15nt)扩增出随机的非相关产物。cDNA 任意扩增产物产生部分 cDNA 片段的文库，ESTs，同样，这些文库可用作基因组作图中的序列标记位点。ESTs 也可以用来分析表达基因的差异(参见差示 PCR)。任意扩增的基因组 DNA 在遗传和物理作图中可用于产生标记。RAPDs 是**随机扩增的多态性 DNA**(randomly amplified polymorphic DNA)标记，在植物基因组作图中广泛应用。随机基因组 PCR 产物用电泳分离后，多态性根据不同的条带很容易鉴定。带有特定的多态性标记的性状共分离是连锁的证据，通过产物分析和发展更特异的引物，RAPDs 可以转变成物理作图中的序列标记位点(见第 12 章基因组与作图)。

随机 PCR 终极形式是扩增出所有的靶序列。达到此目的的一种方法是在所有来源的 DNA 分子上加上接头(linkers，参阅)，用识别接头的引物进行 PCR 杂乱扩增(**连接-衔接 PCR** 或**连接-引发 PCR**，ligation-adaptor PCR, linker-primed PCR)。另一种方法是 DOP-PCR 引物是完全简并的寡核苷酸，这样所有的顺序都能被识别。

PCR 诱变和增大 引物设计不仅用于扩增序列，还可用于改变序列。这主要有两种用途：点突变的导入(体外突变的一种形式，参见)和利于进一步的操作的 PCR 产物末端的修饰。

在第一种方法(**PCR 诱变**)中,引物设计成与靶顺序不完全配对,选择允许退火进行的条件。扩增将末端引物中的突变引入产物中。如果需要在中间位置突变,突变产物本身可以作为一大引物(**megaprimer**),在原先的模板中延伸。或者合并一对相互重叠的 PCR 反应产物,通过限制性酶切和连接(如果有合适位点)或**末端到末端的重组 PCR** (end-to-end recombination PCR)得到终产物。这两种策略在图 21.2 中进行比较。

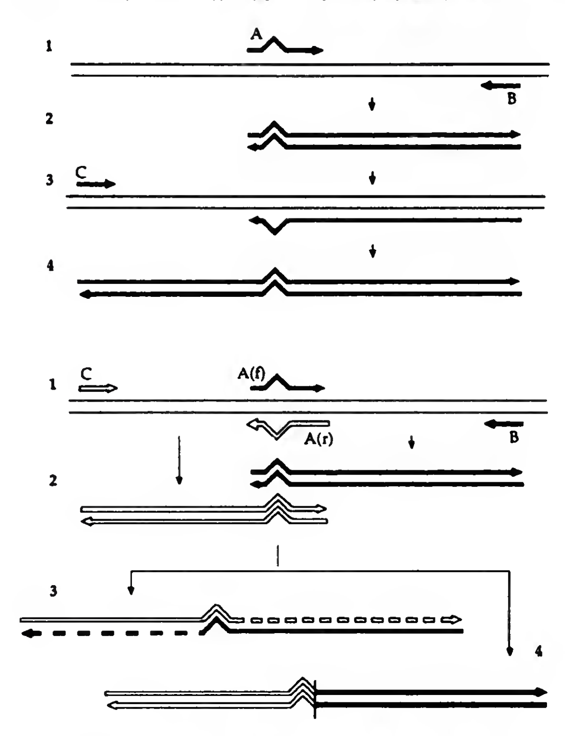


图 21.2 PCR 诱变。可以用不完全配对的引物在 PCR 产物中导入突变,但这只允许在 PCR 产物末端突变。两种策略都允许在靶片段中的任何位置上突变。在大引物的策略中,以错配的引物 A 和 B(1)扩增产生半长产物(2),作为大引物,与引物 C(3)组合产生全长的突变序列(4)。在重组 PCR 策略中,两次扩增(1)产生两个半长产物(2)。两个错配引物在相反方向(A 正向,A 反向),所以产物末端重叠。两个产物混合后通过末端到末端的重组 PCR(3)连接或适宜的限制性酶切位点通过连接(4)得到突变全长产物。

在第二种方法中,引物的 5′端可以加上与模板 DNA 不配对的顺序,使 PCR 产物末端加上额外的顺序。这有很多用途:加上限制性内切酶位点以利于亚克隆,加上通用引物结合位点便于测序和加上噬菌体启动子用于体外转录。

下游应用　当 PCR 产物大量需要并有多种下游应用时,常规的亚克隆技术在体外操作方面还是最方便的。然而 PCR 产物必须以和任何过客 DNA 分子一样的方式(见重组 DNA),插入到克隆载体中。倾向差错热稳定 DNA 聚合酶,如 *Taq* 聚合酶,倾向于以不依赖模板的方式在 PCR 产物的 3′端加上一个核苷酸(通常为 dATP)。而校正酶如 *Pfu* 酶则产生一平端的产物。PCR 产物的亚克隆有多种策略可供选择,参见表 12.1。

表 12.1　PCR 产物克隆的方法

方法	优缺点
平端克隆	所有的 PCR 产物都可以用这种方法克隆，如 *Taq* 酶的产物需要去除突出的碱基，产生平端，亚克隆是没有方向性的。
T-载体	T-载体有突出的 5′胸腺嘧啶能够有效的克隆如 *Taq* 聚合酶有突出的腺嘌呤的产物，产物不需要处理，但 亚克隆没有方向性。
接头引物	引物含有接头，在 PCR 产物末端加上限制性酶切位点。产物在亚克隆前需要酶切，如果 5′和 3′端加上不同的位点，亚克隆是非常有效和有方向性的。在亚克隆前必须确定 PCR 产物内部没有所使用的酶切位点。
DISEC/TRISEC	二核苷酸/三核苷酸黏端克隆。产物和载体需要部分补平和外切酶处理，但亚克隆是有效和有方向性的。

21.2　PCR 基本战略的提高和扩展

反转录 PCR(RT-PCR)　　cDNA 的 PCR 扩增称为 RT-PCR，因为在扩增前涉及反转录步骤。cDNA 第一链的合成反应以常规方式进行(参见 cDNA 合成)，而在 PCR 的第一次循环中进行 cDNA 第二链的合成。有反转录活性的温度稳定的 DNA 聚合酶最近已有报道，在将来有可能使 RT-PCR 在一个反应中完成。

RT-PCR 是扩增 RNA 分子的序列非常灵敏的方法，能够在复杂的来源中检测和分离 cDNA 序列。RT-PCR 作为一种制备技术比起常规文库为基础的 cDNA 克隆方法有很多优越性，包括：快速度，需要实验材料少，能以细胞总 RNA 为材料，不受大量污染 rRNA 和 tRNA 的影响。缺点是倾向于产生部分长度的产物(ESTs)，而且不像文库能作为永久性的实验材料。任意 RT-PCR 是用随机的六碱基或长一些的任意引物和与 mRNA 多聚腺嘌呤尾巴配对的寡聚胸腺嘧啶的引物，产生代表表达情况的标记库以供进一步分析。一种应用是上文提到的，作为基因组物理作图的标记——表达序列标记比非编码 DNA 中所存在的序列更有可能是单一顺序。第二个应用是鉴定差示表达基因。这种技术，被称为**差示 PCR**(different display PCR)或 **mRNA 指纹法**(mRNA fingerprinting)，将不同来源扩增得到的 ESTs 在电泳中的邻近泳道中进行比较。差示表达的基因在电泳上表现为条带的增加或消失，这是非常灵敏的检测被调节基因产物的方法。另一种类似用于鉴定基因组 DNA 差异的方法是根据限制性酶切片段的大小和数量的变化，称为**表现度差示分析**(representational difference analysis，**RDA**)。

作为诊断技术，RT-PCR 用于基因产物的定量和定位。**定量 PCR** 是将靶 cDNA 的扩增与在反应体系中事先加入的已知量的片段进行竞争。通过所扩增的竞争性片段的量作一线性图，来对目标片段定量。**原位 PCR**(*in situ* PCR)是在细胞的自然位置处扩增 mRNA 的 RT-PCR 反应，能够精确地确定很少的转录本。这两种以 PCR 为基础的方法都要比传统的杂交方法(参见 Northern 印迹、RNA 酶保护实验及原位杂交)灵敏。

未知序列的扩增　　基本的 PCR 反应限于扩增已知引物之间的 DNA 序列，这样就需要事先知道序列的情况。然而我们经常需要分析已知序列的 DNA 侧翼区域(未测序)，如根据 ESTs 克隆全长 cDNA 或扩增基因片段的上游调控元件。

反向 PCR[inverse PCR，或**外翻 PCR**(inside out PCR)]用来扩增基因组 DNA 的两侧顺序。如果用限制性内切酶作用 DNA，需要被扩增的序列将位于具有 5′和 3′两侧区域

DNA大片段中。这样的片段经过连接酶作用形成环状的结构,原先用来扩增内部序列的引物就可以用来扩增环状结构的剩余部分的序列。另一些同样目的的策略是在线性DNA分子两侧加上接头,作为引物结合位置,或者利用分子内的二级结构如茎－环结构(**锅柄PCR**, panhandle PCR)和内部的泡结构(**小载体PCR**, vectorette PCR)用于链的特异引发。限制性酶切片段的末端加上接头后,以一已知(基因专一性)引物和接头对应的引物扩增。扩增得到的产物包括了两个引物之间的未知侧端序列。如果基因专一性的引物被生物素标记,产物就可以用链霉抗生物素蛋白捕捉(捕捉PCR, capture PCR),这样那些接头间非特异性的产物就被去除(参见*生物素链霉抗生物素系统*)。一个类似的策略用于从ESTs产生全长cDNA。在这种称为**快速扩增cDNA末端**(rapid amplification of cDNA ends, RACE)的技术中,用基因专一性引物和与多聚腺嘌呤尾巴结合的寡聚胸腺嘧啶引物,扩增cDNA的3′端顺序(3′ RACE)。同样5′端的扩增也可以同样的方式用末端转移酶在cDNA第二链加上人工的尾巴,或者类似捕捉PCR加上一接头,称为**5′ RACE**。

不对称PCR(asymmetric PCR)　　在上文讨论的PCR中,引物对都是以等量的方式产生同样量目标链,从而形成大量的双链扩增产物。另一种标准PCR方法的演变是**不对称PCR**(asymmetric PCR)或**单链PCR**(single-stranded, ssPCR),引物对中一条引物大大过量,在一定正常PCR循环后,另一条引物被消耗完,反应变成一条链的线性增加。这种技术有很多应用,包括产生用于构象电泳分析(参见*突变检测*)的单链DNA;探针合成;序列测定。DNA序列测定可以通过单链PCR,只使用一条引物,并加入双脱氧核苷酸三磷酸(参见*循环测序*)。

21.3　体外扩增的其他方法

其他扩增系统　　PCR是体外扩增用得最广泛的方法,但不是唯一的方法,表21.2中列举了其他一些技术和它们的应用。

表21.2　其他一些体外应用方法

系统	解释和应用
连接酶链反应(ligase chain reaction, LCR)	用于灵敏地检测DNA序列的变化和等位基因的差异。用两个引物在目标基因的邻近位置退火。如果目标序列出现,引物被连接酶连接,可以作为进一步退火和连接的位点,导致引物连接产物扩增。在目标基因缺乏的情况下,没有连接和扩增发生。选择只有完全配对的引物才能退火的条件,能够灵敏地检测等位基因点突变的差异,称为**等位基因专一性连接**(allele-specific ligation),参见*等位基因专一性杂交*(allele-specific hybridization)和*挂锁探针*(padlock probes)。
转录为基础的扩增;(transcription-based amplification);核酸序列为基础的扩增(nucleic acid sequence-based amplification, NASBA)	以转录和反转录为基础的快速扩增方法,靶DNA或RNA通过用含噬菌体启动子的引物的标准技术来扩增。进一步的扩增是用噬菌体RNA聚合酶转录来完成,转录本反转录产生cDNA以供进一步扩增。转录反应比起PCR每循环两倍的增加,可以每次扩增一千倍。
链置换反应(strand displacement reaction)	与PCR反应类似的一种技术,但每个循环不需要变性步骤。延伸后,引物从延伸产物上通过限制性内切酶处理分离出来。进一步的延伸通过缺口处的链置换进行。扩增得到的DNA必须受到保护,以免被剪切。

(陈　淳 译)

第22章
蛋白质:结构,功能与进化

基本概念与定义

- 蛋白质是由一条或多条多肽链组成的大分子,每一条多肽链都是由一系列氨基酸残基通过肽键一个一个连接而成。多肽链的**一级结构**(primary structure)指由遗传密码决定的氨基酸残基的线状排列,这些残基都来自21种基本氨基酸及其衍生物(见第11章遗传密码)。
- 一旦被插入一多肽链,个别的氨基酸就可能会发生**共翻译修饰**(cotranslation modification)或**翻译后修饰**(posttranslational modification)。这可能包括简单的化学修饰(例如甲基化,羟基化),或大的化学基团的添加(例如糖基化,酰化)。较小的修饰可能是永久的并且对蛋白质的正确折叠与功能是必须的(例如胶原蛋白中脯氨酸与赖氨酸的羟基化);但更常见的是可逆的较小的修饰对蛋白质的活性进行调节(例如在真核信号蛋白质中丝氨酸,苏氨酸与酪氨酸的磷酸化;见信号传导)。较大的化学加合物的添加往往是永久性修饰,并且关系到胞内蛋白质的运输与加工(例如很多分泌蛋白质被糖基化的)或者一蛋白质与其辅基共价的连接(例如细胞色素C与血红素基团连接)。
- 一新合成的多肽链必须折叠成它的**天然构象**(native conformation,生物活性的构象)。在天然蛋白质中观察到结构组织的几个不同水平。**二级结构**(secondary structure)指残基局部构型的重复,例如α-螺旋与β-折叠,主要由多肽键单位之间形成氢键的需要而产生。二级结构往往聚集在一起形成更大更复杂的排列(如螺旋-转角-螺旋基序),它们可能会在与其他胞内组分的相互作用中起特定的作用。**三级结构**(tertiary structure)指多肽链的整体三维构型,反映了很多不同类型的共价或非共价的化学键作用,稳定了能量上最适合的折叠构象。**四级结构**(quaternary structure)指多亚基蛋白质中多肽链亚基的排列。考虑到蛋白质可能采取的其他构象(变性构象)的数目十分庞大,人们仍不清楚蛋白质是如何折叠成天然构象的。折叠可由内在因素指导(由一成核中心或一系列适宜的中间体),也可被外在因素指导[如被称为**分子伴侣**(molecular chaperones)的蛋白质]。很多蛋白质可以不同的稳定构象存在并显示不同的活性。这些状态之间的转换可以通过共价修饰或与其他蛋白质的非共价相互作用调节(**别构性**,allostery)。
- 可以基于蛋白质结构与功能的相似将其划分为家族。在经典的家族中,相关蛋白质通过**基因重复**(gene duplication)和**趋异**(divergence)而进化:一祖先基因重复,

等位拷贝由于积累的突变而趋异。突变可以引起序列，结构以及表达模式的变化，导致新的功能的出现。在一些串联排列的基因家族中，由非等位重组事件即不等位交换和/或基因转变引起的趋异的突变压力由**序列均一化**(sequence homogenization)来计算。更复杂的进化方式包括**嵌合蛋白质**(chimeric proteins)，由已存在的蛋白质的**组件**(module，连续的功能片断)重组得到。嵌合蛋白质的进化由于真核基因的外显子/内含子的结构而可能发生：单个的外显子可能发生串联重复，导致单个多肽链中有重复的蛋白质组件[**外显子重复**(exon repetition)]，另外，外显子可能扩散而重复，并克隆其他的基因产生嵌合蛋白质[**外显子改组**(exon shuffling)]。某些蛋白质组件，如锌指 DNA 结合组件，分布十分广泛。

- 蛋白质在最显著的水平表现了基因的功能。但蛋白质结构的解析虽已变得常规化，对蛋白质功能的确定却往往不是直接的。从理论上说，蛋白质的功能可以通过对结构，表达模式，突变效应及与其他细胞组分的相互作用的分析来确定。有时候，一种策略就可以得到目的，但如用一种方法得不到所需的信息时，往往就需要多种策略相互补充(例如由于遗传丰余突变不显示效应)。现在还不能从蛋白质的一级序列信息来预测蛋白质的三级结构，但由于可得到的一级序列不断增多，很多蛋白质的结构与功能可以通过与已经研究过的蛋白质的同源性来了解。着眼于单个蛋白质的深入分析可能会在不远的将来被着眼于整个基因组功能的分析所超越(**功能基因组**，functional genomics)，这包括对胞内合成的所有蛋白质的结构，表达，突变与相互作用的系统性考察。

22.1 蛋白质一级结构

多肽链结构　　**多肽链**(polypeptides)是氨基酸残基(文框 22.1)通过**肽键**(peptide bond)[通过相邻氨基酸的氨基(—NH_2)与羧基(—COOH)缩合而成的键；图 22.1]共价连接形成的线状长链。多肽链的第一个残基仍保留它的氨基，而最后一个残基仍保留它的羧基。因此多肽链的末端分别被称为 **N 端**(N-terminal)与 **C 端**(C-terminal)。蛋白质的**一级结构**(primary structure)指的是**氨基酸顺序**(amino acid sequence)，常规上从 N 到 C 的方向读出。这和核苷酸的一级序列——即核苷酸顺序是同义的，而且多肽链 **N→C 极性**(polarity)与它同来源的 mRNA 的从 5′→3′的极性(见第 16 章核酸结构)是共线性的。所有的新生肽链都以甲硫氨酸起始(见第 23 章蛋白质合成)，虽然它往往在翻译后被切除。在原核与真核生物体中，N 端甲硫氨酸的氨基在与起始 tRNA(参阅)连接后被甲酰化封闭。

多肽链的键构象　　多肽链是由氨基酸单体组成的大分子，那么很容易把它们想象成线上的一串珠子，氨基酸是坚硬的珠子而肽键是柔软的连接物使折叠可以发生。事实上完全相反。肽键采取僵硬的平面式构象，而氨基酸残基中的键，**φ 键**(Cα－N 键)与 **Ψ 键**(Cα－C′键)可以有不同程度的旋转(图 22.1)。肽键(**ω 键**)往往采取反式构象(180°旋转)，但在脯氨酸残基时由于二级氨基基团的环状结构肽键无法旋转。因此邻近脯氨酸残基的肽键采取顺式构象。

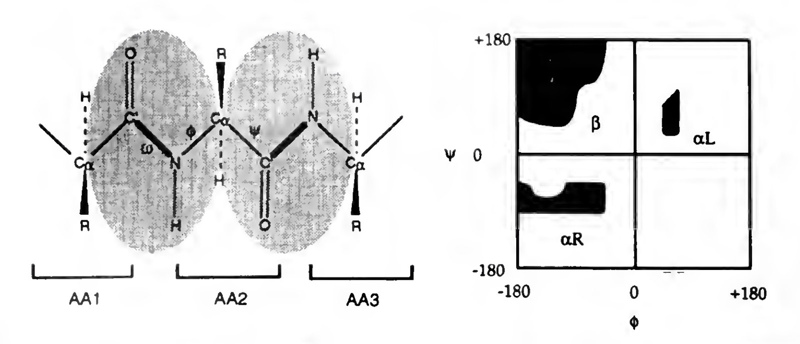


图 22.1 多肽链主链的键构象。由氨酰残基组成的多肽链(在左图中用 AA1,AA2,AA3 表示)由肽键(ω 键)连接而成,用粗线表示。肽键形成平面的刚性的结构(用圆圈界定),往往是反式构象。氨基酸的 ϕ 键与 Ψ 键可以旋转使蛋白质折叠。允许的 ϕ 键与 Ψ 键的范围用 Ramachandran 构象图(右图)表示,其中 β 为 β 折叠,αR 代表右手 α 螺旋以及 αL 代表左手 α 螺旋。Cα 是中心的四价的碳原子,C′是每一肽键的羧基碳原子。

Ramachandran 构象图(Ramachandran plot)是以 ϕ 键与 Ψ 键角度作为不同的坐标绘制的点图(图 22.1)。它表示了能量许可即不被多肽链主链与侧链的立体效应所限制的 ϕ 键与 Ψ 键角度的集合。对大部分氨基酸来说,允许的集合落在三个主要的区域,分别对应于 β 折叠,右手 α 螺旋以及左手 α 螺旋(见下文)。甘氨酸带有由一个氢原子组成的较小的侧链,可以采取更多种类的构象,因而在蛋白质结构的多样性中起重要的作用。然而,多肽链主链的整体折叠可以迫使一些氨基酸采取能量不适宜的键构象。

22.2 高级蛋白质结构

二级结构 蛋白质的**二级结构**(secondary structure)是由分子内氢键和其他的弱键相互作用产生的规则重复的局部构象。虽然氢键是由极性氨基酸侧链(如丝氨酸和苏氨酸侧链)形成的,但多肽链主链自身也具有极性,因为每一肽键单位的氨基氮原子可以作为氢键给体。而羧基氧原子可以作为氢键受体。多肽链上肽键单位的周期性的间隔使规律有序的结构得以形成。已识别到三大类二级结构,分别定义为**螺旋**(helix),**折叠**(sheet)与**转角**(turn)。其中 α 螺旋与 β 折叠是最常见的,在连续残基具有相同的 ϕ 键与 Ψ 键角度时发生(图 22.2)。不同于这三类的二级结构统称为**卷曲**(coil)。

在多肽链中 **α 螺旋**往往是由 L-氨基酸组成的右手螺旋,在连续残基具有相同的分别约为 $-60°$ 与 $-45°$ 的 ϕ 键与 Ψ 键角度时发生,对应于在 Ramachandran 构象图中左下方的四分之一区域中封闭曲线的值(图 22.1)。主链的卷曲使每隔四个残基的肽键单位之间可以产生氢键。α 螺旋中肽键的排列是同向的,这样就放大了肽键单位的轻微极性的电荷分布,使 α 螺旋产生一显著的偶极矩。螺旋之间的相互作用不仅仅反映了个别残基的化学基团的相互作用,还反映了整体结构的累积电荷分布。α 螺旋的长度从 4 残基到超过 40 个残基(1~12 圈螺旋)。在螺旋中出现的氨基酸有一定的倾向性,如疏水残基丙氨酸与亮氨酸经常出现而另一些氨基酸如丝氨酸就很少出现。**两亲 α 螺旋**(amphipathic α helix)一面具有疏水残基相对的一面有极性的或带电荷的残基。这样产生了既可以与极性化学环境也可以与非极性化学环境相互作用的结构(参见亮氨酸拉链,离子通道)。

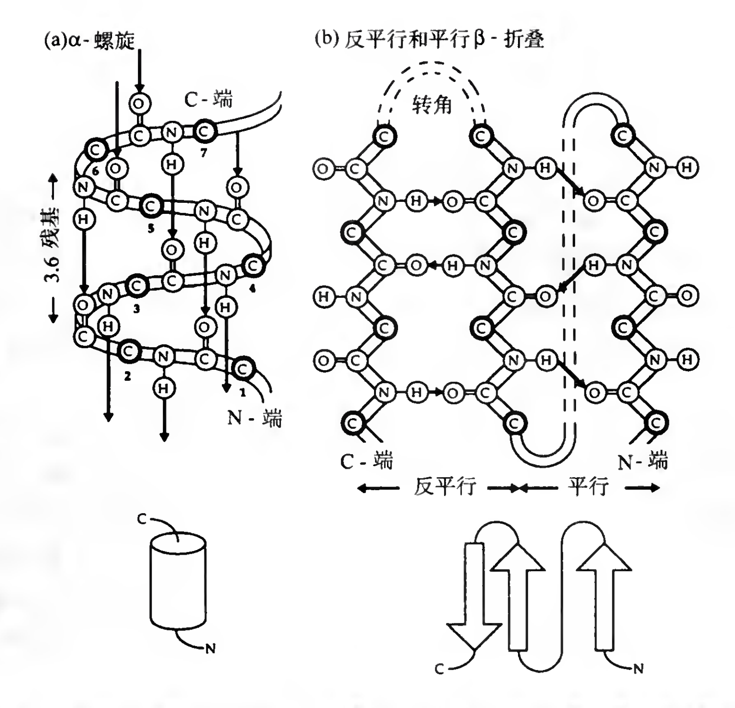


图 22.2　蛋白质中主要的二级结构。(a) α 螺旋。相隔四个残基的氨基氮与羧基氧形成氢键产生右手螺旋，每 3.6 个残基为一周期。箭头表明氢键并指向偶极矩的负极。(b) β 折叠。右手的 β 层片可以通过层片间的氢键(箭头)形成反平行或平行的 β 折叠。反平行的 β 折叠往往被转角分开，而平行的 β 折叠往往被 α 螺旋分开。粗线圈出的碳原子是 Cα 碳原子。在蛋白质剖面图中 α 螺旋与 β 折叠往往分别用圆柱(或卷曲)和箭头来表示。

β 折叠(β-sheet)由多肽链中 **β 链**(β-strands)区域形成。在 β 链中，键角几乎完全伸展，对应于在 Ramachandran 构象图中左上方的四分之一区域中较广泛区域中的值(图 22.1)。几个 β 链以平行或反平行或混合方式排列，使不同 β 链的肽键单位可以形成氢键。这种结构常常被称为**褶状折叠**(pleated sheets)，因为每一链中的 Cα 碳原子是在折叠平面或上或下排列的。β 折叠往往以右手方向弯曲，有时弯曲达到形成类螺旋桨结构的程度。在有些蛋白质中，β 链以相互垂直方式排列以产生篮状排列的 β 折叠类型。

α 螺旋与 β 折叠通过连接残基相连，这些连接残基也可通过氢键产生自身的二级结构，被称为**转角**(turns)。最简单的转角很短，例如 β 转角是由相隔三个残基的肽键单位之间的氢键产生的，导致了多肽链主链形成一发夹状转角。也有可能，连接残基没有形成任何二级结构，这时它们被称为**环**(loops)。由于缺少分子内的氢键，环常与水形成氢键，因而往往在蛋白质的表面被发现。

超二级结构　几乎所有的蛋白质都包含这样的区域，它们采取 α 螺旋或 β 折叠的结构，由转角或环连接。两个以上的二级结构往往组合形成一更加复杂的结构，可以被称为**超二级结构**(supersecondary structure)或**基序**(motifs，又译“模体”)[1]。有些组合是特别普遍的并构成结构性结构域的基础(见下文)或可与胞内的其他分子的发生特异的相互作用。

1) 基序是描述被识别的短多肽顺序的另一种方式，因为它在两种或更多的蛋白中保守。基序的含义并不是仅指上述的结构基序(见本章后面的基序、组件和结构域部分)。

在三种或更多没有关系的蛋白质中观察到的超二级结构被称为**超折叠**(superfold)。已被描绘的超二级结构数目众多,在表 22.1 中列出了有些常见的例子。它们由相邻的 α 螺旋和 β 折叠的侧链相互堆积而成。

三级,四级以及活性结构 蛋白质的**三级结构**(tertiary structure)指多肽链的整体三维结构,反映了二级结构与超二级结构的堆积形成紧密的球状结构域。**结构域**(domain)是蛋白质三级结构的最小单位(见组件),可能可被认为是独立或准独立的功能单位。小的蛋白质可能只包括单个结构域,而大蛋白质一般由多个结构域集合而成,这些结构域在理论上可以独立地或在一嵌合蛋白质中折叠与执行功能(参见*折叠元*,*嵌合蛋白*,*结构域交换*)。

很多多肽链是单独起作用的,而有些则聚集成寡聚与多聚复合物,它们可能包含同一多肽链的多个拷贝[**同聚多亚基蛋白**(homomultimers),例如 β 乳糖苷酶],也可能包括不同的多肽[**异聚多亚基蛋白**(hetermultimers),如血红蛋白]。在多聚复合蛋白中人们认识了更高级的结构层次,**四级结构**(quaternary structure),它反映了单个成分(**原聚体**,protomers)的空间组织以及它们之间形成的化学键。单个的原聚体的功能可能独立或依赖于整个复合物。在后一种情况中,四级结构在蛋白质活性的调节中十分重要(参见*协同结合*)。

当蛋白质与非蛋白因子如辅因子、配体、底物等等相互作用时会形成进一步的结构层次。只有在非蛋白辅因子存在时才能行使功能的蛋白质被称为**结合蛋白质**(conjugated proteins)。它们与它们的辅因子(例如辅酶,核苷酸,金属离子)往往是非共价结合的,在辅因子不存在时它们被称为**脱辅基蛋白质**(apoprotein)而与辅因子结合后被称为**全蛋白质**(holoprotein)。然而也有一些辅因子是与蛋白质共价结合的,在这种情况下被称为**辅基**(prosthetic group,例如在血红蛋白中的血红素)。在两种情况下,蛋白的活性在与辅因子结合后会发生变化,可能是由于辅基对蛋白结构的影响或是由于辅基的特异的化学性质。蛋白质与配体或底物的相互作用可能包括共价与非共价键,虽然前者往往是可逆的(参见*自杀酶*)。配体或底物结合引起的蛋白质结构改变对蛋白质的功能可能十分重要(参见*异构*,*诱导偶合*)。

二级结构的形成主要受氢键影响,但三级、四级与活性结构包括很多不同类型的化学键(共价与非共价)的贡献。由于这些键的复杂性,现在还不可能从一级氨基酸序列预测高级有序的结构。蛋白质的三级、四级结构以及与辅因子、配体、底物的相互作用的结构只能用生物物理的方法来研究(文框 22.2)。

高级蛋白质结构中的共价键 在有些蛋白质中,只有多肽链主链与氨基酸残基基团中存在共价键。然而,很多蛋白质需要另外的共价相互作用以到达它们的天然构象,它们需要半胱氨酸残基的**巯基**(sulfydryl group, −SH)之间形成的**二硫键**(disulfide bonds)。在很多蛋白质中分子内的二硫键是稳定的三级结构所必需的,而分子间的二硫键还可以稳定四级结构。这包括同一多肽链的拷贝之间或是不同基因编码的多肽链之间形成的二硫键(例如在抗体合成中,二硫键连接两条抗体重链,也连接重链与轻链)。胰岛素提供了一个有趣的变异的例子。在成熟蛋白质中有三对二硫键,一对是 A 链的分子内二硫键,两对是连接 A 链与 B 链的分子间二硫键。然而在新生肽链中,三对二硫键都是分子内二硫键

因为A链与B链是由同一基因编码,开始它们之间有一段C多肽连接,在二硫键形成后C多肽被切除。

表22.1 蛋白质中常见的超二级结构的一些例子。它们中间很多可以在SCOP数据库查看(见本章结尾)

结构	组成
简单超二级结构:α螺旋	
卷曲螺旋	两个以上的α螺旋互相卷曲,这种结构特别稳定,往往在纤维结构蛋白质如α角蛋白,胶原蛋白及纤维蛋白原中发现。亮氨酸拉链(参阅)的二聚形成一个两螺旋的卷曲螺旋。普通的α螺旋的螺旋重复是3.6个残基。但在一两螺旋的卷曲螺旋中,扭曲使周期减少到3.5个残基,这样允许每隔七个出现的亮氨酸可以相互作用(见第17章核酸结合蛋白)。
螺旋-转角-螺旋	两个α螺旋由一转角连接。这个结构是一常见的原核与真核的DNA结合基序(见核酸结合蛋白)。
螺旋-环-螺旋	两个α螺旋由一环连接区域连接,是一种常见的基序,在一些DNA结合蛋白质(例如Achaeta-scute,无姐妹(Sisterless-a),肌浆蛋白D(MyoD))的二聚组件中及它们的抑制剂(例如Id,Deadpan)中(参见基本螺旋-环-螺旋,性别决定)以及钙调蛋白中发现。在钙调蛋白中,螺旋E和F采取一种相互垂直的螺旋-环-螺旋结构被称为**EF手**(EF-hand)。
简单超二级结构:β折叠	
β-发夹	由环连接的两个反平行的β链。在很多蛋白质中单独或作为大的β折叠的一部分被发现。没有特殊的功能与这种结构相关。
β-α-β基序	由α螺旋分开的两个平行的β链。大部分包括平行的β链的β折叠使用这种基序,这样β链的相连的两端在β折叠的相对的两面。
β-弓	由环连接的两个相邻的β链形成的结构,但这些层片位于不同的β折叠中。
β-凸起	β链中特殊的残基引起β折叠的扭曲。
复杂超二级结构:α螺旋	
四螺旋束	四个反平行的α螺旋相互堆积形成疏水核,而亲水表面暴露。在很多主要由α螺旋组成的蛋白质如Rop和铁蛋白中发现。
珠蛋白折叠	八个α螺旋以一种复杂的方式排列,这样多肽链一级结构相邻的螺旋在二级结构中不一定相邻。螺旋形成口袋状是蛋白质的活性位点。珠蛋白折叠在进化中高度保守:在血红蛋白与肌红蛋白中它与血红素基团结合。
复杂超二级结构:β折叠	
β-桶	由大的反平行β折叠形成的结构,整个折叠卷曲使第一个β链与最后一个β链通过氢键相互作用从而形成一个近似的圆柱形。每一β层片通过一发夹转角或环与其他层片相连。这种结构往往形成与小分子结合的口袋,例如视黄醛结合蛋白包含与视黄醛分子结合的β-桶结构。
β-螺旋浆	β链扭曲成旋转排列的β-折叠;流感病毒神经氨酸酶蛋白包括6个β-螺旋浆基序,每个由4个反平行β-链组成。这6个基序排列形成一种类桶状结构域。
β-三明治结构	两个以上的β折叠,一个堆积在另一个上面。
希腊钥匙	四链的反平行β折叠。一级序列顺序为(1-2-3-4)的四个链在剖面图上顺序为(4-1-2-3)。在很多具有反平行β折叠的蛋白质中发现。
果冻卷	四个希腊钥匙基序形成一桶状结构,它的名称是由于多肽链主链缠绕在桶状结构上的方式。在流感病毒的血凝素等蛋白中发现。
复杂超二级结构:α螺旋与β折叠	
αβ桶	由顺序的β-α-β基序形成的结构,它们卷曲形成圆柱体,β链是平行的,α螺旋使之封闭。在酶如丙酮酸激酶,醛缩酶,烯醇化酶,RuBisCo,葡萄糖异构酶以及三磷酸异构酶(因此往往被称为**TIM桶,TIM barrel**)等等中常见的基序。
αβ开放折叠	由顺序的β-α-β基序形成的结构,它们没有发生卷曲而是形成折叠,平行的β链的两侧被α螺旋填充。是在很多蛋白质如已糖激酶和磷酸甘油酸变位酶中发现的高度可变的基序。
Rossman折叠	另一类β-α基序,包括中心部分的β折叠及围绕的α螺旋。在核酸结合蛋白与核糖体蛋白中常见。
其他结构	
欧米茄环	长的环,两端靠近。它不是严格意义上的二级结构,因为没有环内的氢键相互作用。

有时辅基与蛋白质之间也有共价键连接。例如细胞色素 C 的血红素基团就与蛋白的几个残基共价连接。在金属离子的结合中往往是配位键在起作用，如在锌指（参见此条）基序中及细菌铜和汞离子结合蛋白质中。但血红素中的铁原子与它的所有的结合蛋白质都是共价连接的，包括细胞色素 C 与血红蛋白。

高级蛋白质结构中的非共价键　　氨基酸可以粗略地分为具疏水（非极性）侧链与具亲水（极性或带电荷）侧链两种（文框 22.1）。在蛋白质中，将显著疏水的残基暴露于水分子是热力学上不能接受的，因而可溶性蛋白具有一**疏水核心**（hydrophobic core）将非极性残基隐藏起来，而极性与带电荷的表面暴露于溶剂。

极少有带电荷的残基被**包埋**（bury）在蛋白质的内部，但极性残基不一定被排斥，而且多肽链主链本身具有极性也穿过疏水核心。极性原子一般与水产生氢键，当被包埋在蛋白质的疏水核心中无法与水作用时，这种产生氢键的趋势必需通过二级结构来承担。另外蛋白质可能被变性——其典型的稳定自由能等同于产生一个或两个氢键的能量。二级结构的形成，主要是 α 螺旋与 β 折叠，中和了主链上的极性原子，极性侧链也可以通过与主链或彼此之间的氢键来中和。疏水核心由范德华力进一步稳定，范德华力（疏水吸引）在中性原子接触之前随它们的距离接近而上升。这由不同二级结构元件的表面互补造成的蛋白质核心原子的紧密堆积反映出来。

蛋白质的表面富含极性与带电荷残基虽然非极性残基也可能暴露出来。以这种方式，蛋白质可以通过静电力、氢键、范德华力与其他分子发生非共价的相互作用。如蛋白质与靶分子带有不同电荷，静电力就起很重要的作用，如在高碱性的组蛋白与 DNA 的负电荷的磷酸主链之间的“盐桥”相互作用。范德华力在互补表面的相互作用中特别重要，这时水分子被排斥，结合使适宜的化学基团十分靠近。相反，水分子可以作为桥梁在氢键等蛋白质相互作用中起重要作用，特别是在蛋白质与 DNA 的相互作用中（见第 17 章核酸结合蛋白）。

蛋白质折叠　　多肽链线状合成后（见第 23 章蛋白质合成），必需折叠以获得正确的二级与三级构象。正确的结构是分子的天然态即生物活性状态。非天然的变性构象数目庞大，并且 **Levinthal 假定**（Levinthal paradox）表明了对所有构象的随机搜索要求非常长的时间。因此，**蛋白质折叠**（protein folding）必需由在能量上有利的相互作用指导遵从－限定途径进行。人们提出了三种模型来解释**蛋白质折叠问题**（protein folding problem）即使蛋白质获得了它们的天然构象过程的总和。

其中两种模型认为折叠过程中有稳定中间态的存在，通过它折叠得以进行。在**框架模型**（framework model）中，多肽链上邻近的残基在没有确定三级结构存在时相互作用形成二级结构。二级结构的可能排列数目有限，呈现出来的这些结构的随机融合足以辨别与达到它的天然构象。**疏水内陷模型**（hydrophobic collapse model）认为非极性侧链将水分子排斥这一在能量上有利的过程起始了形成疏水蛋白质核心的折叠反应，蛋白质的其他部分可以在较有限的可能的构象中进行选择重新组织。在两种情况下都有一较迟出现的紧密堆积的中间体，**熔球态**（molten globule），它富含二级结构并且排列成大致正确的构象但松散堆积。蛋白质折叠的巩固阶段涉及中间态的重新组织以产生天然态。

另外,**成核模型**(nucleation model)认为蛋白质折叠通过多肽链上特殊残基形成的核起始折叠,其他结构围绕着核产生。这个球状折叠的过程不需要中间体,通过一较确定或较不确定的核发生,各种结构围绕着核产生同时核发生缩聚(**核-缩聚**,nucleation-condensation)。很多对蛋白质折叠的研究已辨明了中间折叠状态从而支持了框架与内陷模型。然而最近一些小蛋白质表现出以简单的两态模型折叠,过程中没有中间体,只有单个的过渡态。这种球状折叠单位被称为**折叠元**(foldons)。因此,对很多蛋白质来说最明显的折叠机理是由一系列有效的动力学中间状态导致的类天然的熔球态的形成,而另一些则通过成核机理一步完成折叠。

分子伴侣 在体内,很多新合成的蛋白质不能自发地折叠成它们的天然构象(**自组装**,self-assembly),可能是由于折叠过程要求能量上不利的中间态的过渡或由于同时存在几种可能的稳定折叠途径,只有一种可以达到天然态。在这种情况下,正确的折叠由被称为**分子伴侣**(molecular chaperones)的分子来指导,它们可以结合在天然态包埋在蛋白质内部而在变性态暴露在蛋白质外的残基(暴露于溶剂的疏水残基)来识别变性态。正确的折叠反映了分子伴侣-底物结合与释放的一个周期,一般依赖于 ATP 的水解以及分子伴侣附属蛋白质的活性(**共分子伴侣**,cochaperones)。分子伴侣对防止胞内变性蛋白质之间的错误相互作用是必须的,而这种错误的相互作用可能会导致如积聚等后果(参见包含体,朊病毒)。

分子伴侣在很多水平指导蛋白质折叠,去折叠,重折叠与聚合,包括①蛋白质合成后起始折叠;②热诱导变性等过程后的重折叠(很多热休克蛋白质与其他应激诱导蛋白质是分子伴侣);③越膜过程中的去折叠与重折叠;④异构蛋白质的不同构象之间的转换;⑤重折叠,如作为酶活性的一部分的变性过程之后;⑥蛋白质降解的准备;⑦多聚体与蛋白质-配体复合物的形成的控制。

自然界有很多分子伴侣家族,它们的特异性与折叠机理不同。**核质蛋白**(nucleoplasmins)是核分子伴侣,控制核小体的聚集(见第 3 章染色质)。另外的重要的分子伴侣家族包括 **Hsp70** 家族,可以稳定新生肽链和利于膜转位,还有**陪伴蛋白**(chaperonins),控制折叠的起始。随着 *E. coli* DnaK(Hsp70 家族的一个成员,与两个共分子伴侣 DnaJ 与 GrpE 共同作用,要求底物的结合与释放)的结构的解析与在 *E. Coli* GroEL(原始型分子伴侣)的折叠循环中分子伴侣与其底物的结构改变的确定,分子伴侣活性的机制变得更清楚了。GroEL 形成一 14 聚结构包括两个内翻的七元环,一端由一共分子伴侣 GroES 的七聚体盖住。每一 GroEL 亚基结合 ATP 以提供底物结合的能量。GroES 促进正确折叠产物的释放。全部的折叠过程要求好几轮 ATP 水解。

另外,研究得不很透彻的分子伴侣家族包括 Hsp90 和 Hsp100,主要在蛋白合成后起作用,可能在应激过程之后防止和/或逆转蛋白质错折叠与积聚。还有,当有些分子伴侣主要是在胞质中起作用,另一些,如原核生物中由 *E. coli sec* 基因编码的 Grp78,特异地在分泌系统中起作用(文框 22.3)。

别构性和协同性 不同的**构象体**(confomers,构象异构体)或不同的四级结构之间的转换常常可以用来调节蛋白质活性。蛋白质构象通过与其他分子的相互作用来控制,还可能涉及可逆的翻译后共价修饰或非共价的聚集。这几种情况都导致蛋白质功能的可逆转

换。

通过共价修饰调节蛋白质活性在原核信号传导与基因表达的控制中很常见。很多蛋白质通过磷酸化调节，例如细胞内信号蛋白如 MAP 激酶，及如视网膜母细胞瘤蛋白 RB1 这样的转录因子(见第 28 章信号传导，第 2 章细胞周期)。另外，组蛋白的可逆乙酰化在染色质结构的控制中很重要(见第 3 章染色质)。共价修饰以两种方式控制蛋白质活性。首先，它们可以影响蛋白质的三级结构以及四级相互作用和与其他分子的相互作用，例如，通过暴露或包埋一特定区域如一催化位点。第二，修饰可能直接在相互作用中起作用。例如酪氨酸的磷酸化使蛋白质添加了负电荷，引起了三级结构的补偿性的重组织。然而，磷酸化的酪氨酸也可作为某些配体如包含 SH2 结构域(参阅)的蛋白质等的直接结合位点。

非共价的相互作用对蛋白质构象的调节包括当配体结合于某一位点时会诱导另一位点的构象改变。蛋白质可能具有几个活性中心，通过构象改变相互联系，这种现象被称为**别构性**(allostery)。配体可能是一小的效应分子(例如在转录调控中——参见转录调控，lac 操纵子，核受体家族)、另一种蛋白质(例如在朊病毒疾病的传播中，见下文)或一细胞内信号蛋白(例如在生长因子对受体酪氨酸激酶的激活中，或在配体控制的离子通道的开放中；见信号传导)。朊病毒可能代表了一异构控制链反应的特殊例子：朊病毒是一正常细胞蛋白质 PrP 的病理构象异构体，朊病毒与正常 PrP 的接触可能会诱导后者产生构象改变，导致它转化为朊病毒(参见重折叠模型)。由于是一种错折叠的构象，朊病毒会发生聚合，在细胞内积聚形成病理性斑点(参见传染性海绵状脑病)。

蛋白质活性的构象控制的另一个例子是**协同结合**(cooperative binding)。在这种情况下，调节发生在四级结构水平：与配体结合后原聚体(protomer)改变构象并且通过四级相互作用提高其他原聚体的亲和力。血红蛋白与氧协同性结合：氧气与一珠蛋白的结合引起其他珠蛋白的构象改变，提高了它们的氧结合效率。某些蛋白质相互协同性结合。单链 DNA 结合蛋白质(SSB)与 DNA 的结合引起蛋白质的构象改变，使其他 SSB 分子以更高效率与之结合，引起围绕着 DNA 的蛋白质丝状体的形成。

22.3 蛋白质修饰

蛋白质修饰的类别　在合成过程中或之后，在多肽链成为功能性蛋白质之前，它们需经历某些形式的共价修饰(表 22.2)。从结构上看，这些修饰可以被分为三种类型：①微小的取代——氨基酸侧链的微小改变；②增大——特定氨基酸残基上添加体积较大的基团；③剪切——从新生肽链上除去残基。从功能上看，修饰可分为蛋白质功能中性或功能必需的，或蛋白质折叠、运输和加工所要求的。很多增大是蛋白质的分选所必需的(一些是蛋白质功能必需的)。几类修饰在蛋白质周转与降解中起重要作用。

修饰还可分为永久与可逆的。剪切与很多增大是永久的。取代可能是永久的，具有重要的结构上的作用(例如半胱氨酸残基的二硫键的形成)，也可能是可逆的，行使调节功能(例如酪氨酸残基的磷酸化)。体内存在细胞编程的酶促修饰，也有另一些修饰以非酶促方式发生，往往与蛋白质寿命有关。这些修饰包括氧化、脱氨基化和在血蛋白质中，与葡萄糖的反应。

表 22.2 程序化酶促蛋白质修饰在蛋白质结构与功能,蛋白质导向或加工以及遗传信息的流向中的作用简表

共价修饰	例子
取代(较小侧链修饰)	
较小侧链修饰——永久性并与蛋白质功能相关	在胶原蛋白中脯氨酸残基的羟基化稳定了三股螺旋组成的卷曲螺旋的三级结构。 在某些激素中酪氨酸残基的磺化。 甲状腺球蛋白的碘化。 凝血酶原谷氨酰胺残基的 γ-羧化。
分子内或分子间键的形成	很多细胞外蛋白质中二硫键的形成,例如胰岛素,免疫球蛋白。
较小侧链修饰——可逆性,与活性调节有关。	酪氨酸、丝氨酸和苏氨酸的磷酸化调节酶的活性。例如受体酪氨酸激酶,细胞周期蛋白激酶(参阅)。 很多侧链发生甲基化,这种修饰的功能仍不清楚。 组蛋白(参阅)赖氨酰基的乙酰化调节它们形成高级染色质结构的活性并在染色质结构域(参阅)的建立中起重要作用。
增大(较大的侧链或主链修饰)	
侧链上化学基团的添加——与蛋白质功能有关	酶活力所必需的核苷酸的增加(例如在大肠杆菌中谷氨酰胺合成酶中腺苷酸基团的添加) 在一些真核胞浆蛋白质中丝氨酸或苏氨酸残基上乙腺氨基葡糖的添加 *Hedgehog* 家族(参阅)信号蛋白质上胆固醇的添加控制其扩散。 结合蛋白质上辅基的添加,例如细胞色素 C 与珠蛋白上血红素的添加。
侧链上化学基团的添加——与蛋白质导向与加工有关	半胱氨酸的酰化使蛋白质导向细胞膜。 GPI 膜的添加使蛋白质导向细胞膜。 在Asn-Xaa-Ser/Thr 的顺序中天冬酰氨残基的 N-糖基化是使蛋白质进入分泌途径的常见修饰(文框 22.3) 在高尔基体中丝氨酸或苏氨酸残基的 O-糖基化 蛋白质的泛肽化使蛋白质定向降解。
末端残基修饰	很多胞浆蛋白质的 N 端氨基酸残基的酰化往往与蛋白质的转换有关。 N 端氨基酸残基的酰化使蛋白质定向到细胞膜,例如 Ras 的豆蔻酰化(参阅)
剪切(残基的除去)	
肽键的剪切	在很多胞浆蛋白质起始的甲硫氨酸共翻译或后翻译剪切 分泌型蛋白质在穿越内质网膜的转运过程中信号肽(参阅)的剪切 非成熟蛋白质(**蛋白原**,proproteins)通过剪切成熟:例如**酶原**(zymogens)蛋白水解后活化;胰导素原内部 C-多肽的除去;Hedgehog 蛋白质被剪切成 N 端与 C 端片断。 遗传信息的加工:例如脊髓灰质炎病毒基因组和哺乳类速激肽基因合成的多蛋白(参阅)的剪切,内蛋白子 inteins(参阅)的剪切。

22.4 蛋白质家族

常规的基因与蛋白质家族分子分类学　**基因家族**(gene family)指具有显著的序列同一性的一些基因,**蛋白质家族**(protein family)指具有同样联系的基因产物。基因或蛋白质

家族中具有特别的相似性的成员又可被归类为**亚家族**(subfamily),更远的关系可分别用**超家族**(superfamily)或**巨大家族**(megafamilies)这样的术语来描绘,代表更高级别的分类学。

常规的基因家族在编码区域的全长中序列保守。这样的家族的进化可以两种方式解释——源于同一祖先的趋异或源于无联系的祖先的趋同。蛋白质的平均长度大约是300～350残基,对应于大约1kp的编码区域。因此随机独立趋同到完全相同核苷酸顺序的概率是$1:4^{1000}$,两个序列独立进化到50%的顺序同一性的概率是$1:4^{500}$。因此以中等或较长的物理距离范围内的基因中的序列保守性被认为是**同源性**(homology)[1)]的基本标志性的证据,这个术语用于描绘来源于相同祖先的关系。与之相反,较小的序列基序如转录调节元件和剪切位点中,有显著的可能性通过随机过程独立发生。

序列趋异反映了随着时间流逝突变的积累,如序列不断趋异,在某一时刻就会发生同源性无法通过序列的保守性来监测。通过序列对比可监测到30%的序列同一性,那么这些序列显然是同源的。当序列等同性小于30%,这种较远的同源性必需通过计算机分析确定。同源性可以通过保守的蛋白质折叠(即保守的三级结构)来辨别,但看起来相同的三级结构往往来源于趋同,而一级序列则完全不同。使氨基酸序列转变成三级结构的蛋白质结构密码是高度简并的;因此很多不同的序列可以采取相同的结构。结构相似暗示的同源性是不确定的,除非被序列关系所支持。

基因家族的关系可以被描绘为定向同源与平行进化同源两种。**定向进化同源**(orthologous)关系指种属之间趋异的基因,即不同种属中行使相同功能的家族成员(例如人类β-珠蛋白和鼠β-珠蛋白)。**平行进化同源**(paralogous)关系指同一基因组中重复与趋异的基因。平行进化同源基因是多基因家族成员(参阅)(人类β-珠蛋白与人类α-珠蛋白)。在大的多基因家族中,联系特别紧密的基因可以被归类于**平行进化同源亚群**,例如小鼠*Hox*基因是平行进化同源的,而代表每一簇中相当的位置的*Hox*基因彼此联系较之它们与其他*Hox*基因之间联系更紧密(见第6章发育,分子方面)。

基因重复与趋异:新功能的起源　从简单生物体到复杂生物体的进化提出了一个问题:新功能在何处产生?增加生物的复杂性(例如从单细胞生物转变为后生生物体)要求本质上由基因决定的功能数目的增加。那么,新功能的起源应该反映了新基因的产生,或已存在的基因的使用方式的改变。在进化中两种过程都会发生:生物复杂性上升总的基因数目上升(见基因结构与作图),同时单个基因产生多个产物,特别是通过可变剪切的方式(参阅)。

在理论上新基因可以通过三种来源发生:①从其他基因组输入;②通过已存在的基因

---

1) 严格来说同源性是一个绝对的术语,即两个顺序或者是同源的或者不是-不存在同源程度的差别。为了定量顺序之间的关系使用**同一性百分数**(percent identity,完全保守的碱基/氨基酸的出现频率)或**相似性百分率**(percentage similarity,完全保守或保守改变的氨基酸出现的频率)。(但是嵌合蛋白质可能被描绘成**部分同源**,partially homologous,见正文)。同源顺序之间的联系程度反映了趋异之后经历的时间(参见分子钟)。因此定向进化同源顺序的比较产生了**分子种系发生**(molecular phylogenies)的认识,而平行进化同源顺序的比较提供了基因家族进化的信息。序列对比的一个问题是缺口的导入。序列可能会通过核苷酸或氨基酸的插入或缺失也可能通过取代而趋异,问题是通过导入缺口,任何一对序列都可以相配,所以在序列对比中必需有缺口处罚(gap penalty)这样导入的缺口越多两顺序之间的等同性越低。这往往是一种主观确定的值,对导入新的缺口比已存在的缺口更高。

的内源重复；③ 由非编码区域中点突变的积累随机自发发生。第三种可以忽略不计，因为它不太可能发生（见较早的概率计算）。基因的输入（水平基因转移，horizontal gene transfer）是相对较少发生的事件（见第 10 章细菌中的基因转移；参见 Ti 质粒，混栖 DNA，急性转化型反转录病毒）并且严格来说不一定产生新的功能，虽然它们对被输入者来说是新的。大部分新基因是由内源**基因重复**（gene duplication）接着发生趋异产生的。这种事件的频率解释了细胞内基因组中多基因家族的优势。

一旦基因重复发生，细胞内就有同一基因完全相同的两份拷贝，这些基因开始会显示表观等位（参见假等位基因，基因丰余）。如果基因产物是必需的，选择会作用于其中一个座位，而另一座位可以发生突变（图 22.3）。在很多情况中，突变的积累会导致功能的丧失产生假基因（参阅）并会渐渐完全丢失。偶然地，突变以某种方式改变了基因产物的结构使它可以行使新的功能。然而更常见的是新功能的获得是由于新的表达模式的获得，接着结构趋异。举例说明，已发现后生生物中许多酶都表现为由显著差别表达的基因编码的多种同工型（**同工酶**，isoenzyme，isozyme）。在很多哺乳动物中的糖酵解酶烯醇化酶有三种同工形式，一种广泛分布，另两种分别在成体的神经元与肌肉中特异表达。这可能代表了趋异过程的早期阶段，每一蛋白质承担同样的功能，只是适应于不同的胞内环境例如神经元中烯醇化酶具氯化物耐受性。甚至更明显分化的蛋白质如果蝇蛋白质 Paired 和 Gooseberry 以及小鼠 Pax-3 蛋白质，如果基因处于异源调节元件的启动下（例如 *gooseberry* 在 paired 启动子下表达可以拯救 *paired*$^-$ 突变果蝇），它们也可以互相代替。进化过程可能既受顺式 DNA 元件中的突变推动又受影响蛋白质结构的突变推动。

另一种产生独立功能的更复杂的路线是组合作用，开始丰余基因拷贝趋异但保持相关的功能并可以协同地产生进一步的功能趋异性。例如在人类和其他哺乳动物中祖先 G 蛋白偶联受体基因的广泛重复提供了另外的嗅觉受体的多样性。这些受体单独或复杂地

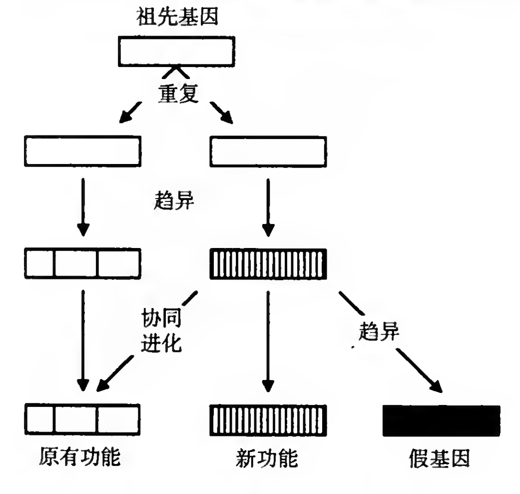


图 22.3 蛋白家族的进化。单个的祖先基因可能重复，在这种情况下选择压力只集中施加于一个拷贝上，而另一个拷贝可能发生突变（带线条的框）。这种歧化可能导致产生新的功能或功能丧失（产生假基因）。在串联基因簇中，通过非等位重组个体序列的趋同（homogenization）可能导致协同进化。

组合成信号网络来辨别不同的味道。

当承担了新的功能，重复的基因就变成非等位的。因此基因重复与趋异的遗传后果开始于完全丰余，进而部分丰余渐渐变成完全功能独立。然而也有可能没有联系的基因趋同产生同一功能，特别在真核生物的复杂信号途径中会发生种种趋同与趋异的交织(见第 28 章信号转导)。一个例子是脊椎的组织者分泌至少三种不相联系的蛋白质——头发生素(Noggin)、Follistatin 和 Chrodin——以通过骨形态发生蛋白质来阻断信号(见第 6 章发育，分子机制)。

基因重复的机制　基因重复，平行进化同源基因的产生可以发生在三个水平：①分离的基因重复事件；②整个基因组的重复；以及③涉及整条染色体或染色体片断的大范围重复。

选择性基因重复(selective gene duplication)事件可能通过一系列不同的机制发生，这取决于已存在的基因的拷贝数。如果在基因组中已有一个以上的拷贝，那么不等交换，不均等姐妹染色单体交换及复制跳格(参阅)都可能在重复 DNA 序列中产生重复(也可能产生缺失)。基因拷贝首次重复是如何发生的？单个基因的重复涉及更加复杂的机制，包括①整个基因组范围的 DNA 复制转座；②染色体断裂后发生的非正常的姐妹染色单体末端连接；以及③可能随机发生的较短的顺序基序的不等交换。涉及大的非连锁 DNA 区域的更为复杂的重排可能对分离基因重复也有贡献(参见基因扩增)。

很多平行进化同源基因的关系可能反映了古代**全基因组重复**(whole genome duplications)事件，特别是在较大的基因组中。基因组大小与基因数目的比较分析表明人类基因组通过至少两轮全基因组重复而产生。很多基因在果蝇与线虫基因组中只出现一次，而在哺乳类中却有四份平行进化同源拷贝，最重要的例子是四个 *Hox* 基因簇(参阅)。全基因组重复在植物中经常发生，而且虽然在哺乳类中是有害的，在很多无脊椎动物中却可被耐受(参见多倍体)。在文昌鱼(*Amphioxous*)的基因组中具有单个的 *Hox* 基因簇，而它被认为是现存与脊椎动物亲缘关系最近的生物。因此，有可能后来进化为脊椎动物的某一古代种族经历了一系列基因组重复事件然后发生染色体水平的趋异以保持二倍体。推测四倍体状态在哺乳类种系中也是瞬时的，因为大尺度的染色体重排将限制同源染色体在减数分裂中与特异的伴侣配对(然而很多鱼是四倍体)。趋异同时发生在基因水平，包括一些基因的丢失与其他基因的重排。最后，只有很少的古代重复事件的痕迹保留下来，如**平行进化染色体片断**(paralogous chromosome segments)。例如，人类染色体 12 与 17 各包含四个 *HOX* 簇中的一个以及几个其他的基因家族(胶原蛋白，烯醇化酶，维甲酸受体，角蛋白，整联蛋白，*WNT* 以及乙醛脱氢酶)的平行进化的成员，这些成员具有较广泛的保守性。

大规模的染色体重排对多基因家族的进化也有贡献。哺乳类基因组的比较图谱(参阅)显示这种重排频繁发生，因而同线区域(syntenic regions，参阅)被限制在较小的染色体片段。染色体重排使四倍体化的证据变得模糊，同时它还导致大规模的**亚基因组重复**(subgenomic duplication)事件。一个特别的例子是人类染色体 1 的两臂似乎是平行进化的，每一臂都包含谷氨酸草酰乙酸转氨酶，凝血因子，两类 tRNA，及铁蛋白重链。这表明染色体 1，人类最大的染色体，可能由古代的包含两个较小的染色体的罗伯逊易位(参阅)事件所产生。

序列趋异与均化 如上面所讨论的，基因重复往往接着发生序列趋异，可能导致新功能的获得或者其中一个拷贝功能丧失。如果重复的基因确实进化产生了新的功能，它们在较晚的物种形成事件中会倾向于保守，导致定向进化同源基因比平行进化同源基因的保守性高很多。在很多分散的多基因家族中都观察到这种情况，例如烯醇化酶等同工酶家族中，哺乳类的定向进化同源基因较之每一物种的平行进化同源基因保守性高很多。

然而，在串联排列的基因家族中，非等位重组事件如不对称交换、不均等姐妹染色单体交换以及基因转变等的高频发生导致了**序列趋同**(sequence homogenization)。在这种情况下，平行进化同源基因较之直向进化同源基因更为相近，因为独立发生的突变在各个物种中可能被固定。这种现象被定义为**协同进化**(concerted evolution)，在串联成簇的基因中普遍发生，如组蛋白基因，但并非所有基因簇都会出现。例如，珠蛋白基因簇，表现出分散基因典型的进化特征(即直向进化同源基因比平行进化同源基因保守性更高)，这可能是因为这些基因在发育中具有不同的作用因而顺序均化可能是有害的。rRNA 基因(参阅)提供了在分散的重复顺序中协同进化发生的特例(在这个例子中，每一个分散的重复顺序由一系列串联排列的 rRNA 基因组成)。这反映了 rDNA 簇在核仁中的紧密缔合，允许频繁的反式非同源重组事件发生以维持基因的均一性(参见卫星缔合)。分散基因之间的反式相互作用是很稀少的事件(参见同源依赖沉默、共抑制、反式感应)。

蛋白质嵌合体：基序，组件和结构域 经典的基因家族在它们的全长顺序中表现出同源性，它们的进化可以用简单的复制与趋异来讨论。然而，基因家族的关系也可能只涉及出现在两个蛋白质中的特定的保守区域，而蛋白质的其他部分完全不同。在一个基因或蛋白质中任何保守顺序单元可被称为**盒**(box)或**基序**[1](motif，又译“模体”)——基序可能是短的顺序元件如 DEAD 盒 RNA 螺旋酶基序，到较大的功能单位如锌指，后者被称为组件可能更合适。**组件**(module)指表现出特异的功能的蛋白质连续片段。组件一般由几个基序组成，虽然锌指是个例外。组件与蛋白质结构域的不同在于后者是三级结构的单元而不一定由一级顺序连续的元件组成。定义的这一差异的重要性在于通过从一个基因中移动连续的 DNA 片段到另一个基因中，可以在蛋白质中插入或删除组件；而这不是对所有结构域都是可能的。不幸的是，在使用中基序、组件与结构域这些术语往往外延模糊，并且常被混用。

部分同源的基因或蛋白质(即在特定的保守片段同源)可能由于对某一特定组件的强烈选择而蛋白质的其他部分可以自由地突变或趋异而发生。在另外的例子中，发现同样的基序与组件出现在来自明显不同谱系的蛋白质中。通过一系列基因重排过程，组件可能以新的方式混合产生新的**嵌合蛋白质**(chimeric protein)。这些使分子分类学复杂化因为蛋白质可能同时包含代表几个不同家族的组件，往往不可能确定每一组件的来源。

组件结构的进化 单个蛋白质的重复组件与不同蛋白质中相关组件的进化包含与整个基因重复(串联或分散)相同的机制，但以亚基因尺度发生。在真核生物中组件的进化是丰富的，因为内含子的出现使编码 DNA 的片段可以在不需要精确的重组连接的情况下

---

1) 基序这个术语是用来以不同的角度描述蛋白质中的超二级结构的单位，如含两个或更多二级结构元件的构型。一个保守序列的基序可能和结构基序是一致的，但并不一定如此，注意盒(box)这个术语不仅是描述基因和蛋白中的基序，而且也描述非编码 DNA 中的基序(如 TATA 盒)。

发生重组。

基因内的重复组件可能是由于**外显子重复**(exon duplication，或**外显子再现**,exon repetition)即单个外显子经历串联重复。这有两个直接的进化益处:①对结构蛋白质来说,外显子重复可以延伸一特异的结构域(例如胶原蛋白中的三螺旋结构域,为多个同样的外显子编码);②对其他蛋白质,外显子重复提高了特异功能结构域的剂量(**剂量重复**,dosage repetition),这样可能提高了蛋白质的活力,或者在特别的例子中,使翻译后剪切能产生多个功能单位(如速激肽和泛素基因)。如同整个基因的重复和趋异,外显子重复往往接着发生**外显子趋异**(exon divergence),产生具有不同的功能而结构保守的组件。重组会引起**外显子缺失**(exon deletion)和外显子重复,使蛋白质中特定的组件被移去。

外显子的分散重复可能在已存在的基因的内含子中引入单个外显子。如果这个外显子对应于一蛋白质组件,接受此外显子的蛋白质可能会被授予新的功能。非等位基因之间的重组会导致这类事件发生,可能在非等位内含子中包含短的重复元件或*转座子*(参阅)。然而除非出现稀少的双交换事件,这更倾向于引起基因融合而非单个基因整合。外显子插入的一更可能的解释是转座因子活性,例如不正常切除(两侧的部分寄主 DNA 代替部分的转座因子被转座),或由于协同转座(两个因子协同作用使其间的 DNA 片段转座)(见第 13 章可移动的遗传因子)。混合已存在的组件产生嵌合蛋白质被称为**外显子移动**(exon shuffling)。它提供了从组件产生新的蛋白的快速途径。

嵌合蛋白进化的进一步的形式为**全基因融合**(whole gene fusion)。例如真核的**多酶蛋白质**(multienzyme protein)在原核中的定向进化同源蛋白就是由不同的基因编码。例如,脊椎动物中的一种多酶蛋白具有三种核苷酸生物合成活性:甘氨酰氨核苷酸合成酶(GARS),氨基咪唑核苷酸合成酶(AIRS)和核苷酸转甲酰酶(GART)。在脊椎动物中这些活性分别由单个 *GART* 基因编码的蛋白中的不同组件来承担。在细菌中,三个酶由三个不同的基因来编码而在酵母中 GARS 与 AIRS 由一个酶承担而 GART 分别编码。在果蝇中情况更为复杂,其多酶蛋白质有四个结构域因为 AIRS 组件被重复。最初曾认为这种全基因融合事件包括一个基因的最后一个内含子与另一个基因的第一个内含子之间的重组。然而在脊椎动物蛋白的分隔 GARS 与 AIRS 组件的外显子中聚腺苷酸位点的发现。暗示了融合可能发生在基因间 DNA,接着调节元件与剪切信号的修饰使共翻译与加工成为可能。

所有形式的外显子重排——重复,移动,缺失和全基因融合——的一个重要方面是**内含子相**(intron phase)(图 22.4)。这涉及编码区域如何被中断因此只应用于编码区内含子。最普遍的内含子是 **0 相**(phase 0),编码序列在密码子之间被打断(即在一个密码子的第三个碱基与下一个密码子的第一个碱基之间)。**1 相**(phase 1)与 **2 相**(phase 2)内含子分别在密码子的第一与第二位或者第二与第三位之间打断。单个内含子的相没有影响,但由外显子重排产生的**融合内含子**(fusion intron)的相在新基因的翻译中极为重要。外显子可根据两侧的两个内含子的相来定义,起始与结束碱基的相对密码子边界的位置相同的外显子被称为**同相外显子**(in-phase intron)。这些外显子可经历无限制的重复并参与外显子移动事件,因为它们不干扰整体的读框。反之,**异相外显子**(out-of-phase intron)开始和结束于密码子内的不同位置(即两侧内含子具有不同的相),这样的外显子插入产生移码。即使同相外显子如通过与一具有不同相的内含子融合插入一基因也会引起

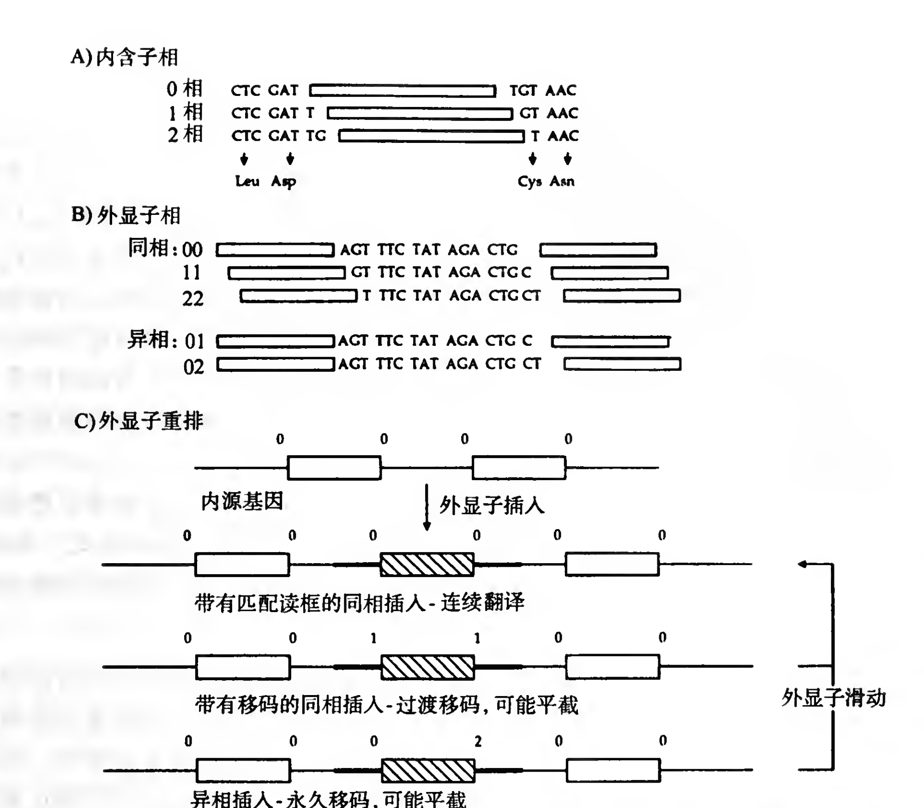


图 22.4 内含子相与外显子重排的后果。(a)内含子相由编码区间断处密码子位置来定义。(b)外显子相由两侧内含子的相定义。共有三类同相外显子与六类异相外显子(显示了两个例子)。(c)具有两个 0,0 相外显子的内源基因。插入一 0,0 相外显子后导致新的外显子的连续翻译。插入一 1,1 相的外显子导致一瞬时的移码,这样插入的外显子的翻译不按正确的读框进行。这往往是有害的,除非这个外显子很小。插入一异相的外显子引起不正确的移码。瞬时与永久的移码都可由小的插入或缺少来纠正,或通过改变外显子与内含子边界的剪接位点突变来校正(外显子滑动)。

内在的瞬时移码。在这些情况下新的外显子的解读是不可预测的,并可能通过出现偶然的终止密码子引起基因的中断。对基因功能的这种改变可以被导致**外显子滑动**(exon sliding)即外显子相对于两侧内含子的起始与终止位置的改变的点突变所校正。这可能由小的插入与缺失或影响剪切位点位置的突变引起。外显子滑动也可以校正异相外显子参与的“滑动事件”。

内含子的进化起源 在较高等的真核生物中,大部分基因为内含子所间断,而在细菌与很多低等真核生物中往往缺少内含子(参见*基因结构*)。自从 1977 年内含子被发现,对其起源与功能一直存在着争论。一种理论是剪接体内含子是由自我剪接内含子获得反式剪切的能力进化而来(内含子起源的自在 DNA 模型)。核内含子与第二类自剪切内含子的剪切机理相同支持这种模型(见第 27 章 RNA 加工)。然而两类内含子识别的剪切位点不同。有可能核剪接体内含子与第二类内含子由同一谱系进化而来。其他模型认为核内含子是分别进化的,是分隔古代基因的基因间 DNA 的副产物,或者在选择性剪接的进化

过程中被放弃的编码片段。后一种模型较为激进,因为它暗示剪接在内含子之前进化,在不存在内含子的情况下某些现代的基因可经历可变的外显子剪接支持这种假设。

另一争论的主题是内含子存在的年代。**早内含子理论**(early intron theories)认为内含子在最早的基因中出现,并在原核谱系被选择性丢失。早内含子可能来源于自我剪切内含子,或进化自分隔古代基因的非编码 DNA,并通过**反向剪接**(reverse splicing)传播。在后一种模型中,剪接的进化受逐渐增大的基因大小所驱动。早内含子理论为古代基因家族中内含子位置的保守性所支持,如珠蛋白基因家族。外显子与蛋白质的功能性组件(据推测为原来基因留下的痕迹)之间的对应关系也支持这种理论。**晚内含子理论**(late intron theories)认为内含子是在相对较近的阶段插入基因的。这种理论为随机内含子的插入所支持,即在基因家族中插入的内含子既不是保守的也不是将蛋白质分隔成整齐的组件,如在胶原蛋白基因家族中发现的内含子。

无论它们的起源是什么,在真核生物的辐射期间内含子快速增生扩散并促进组件进化(内含子重复与移动)和不同的剪接,此可作为增加真核基因组的功能丰度的一种机制。内含子的大小也增加了,因而获得脊椎动物基因不仅富含内含子,而且内含子物质占主要地位(参见基因结构;参考河豚基因组)。

蛋白质家族的选择性扩展　对基因组与 EST 计划(见基因结构与作图)的信息的分析可以证明特定蛋白质组件的分布与丰度。从全基因组测序或很多进行中的基因组测序计划而来的数据显示特定的蛋白质家族非常成功,但这种成功往往是谱系依赖的。很多广泛存在的蛋白质组件包括组成代谢酶与蛋白质合成装置的核心组分的组件表现出最小程度的增生扩散。反之,其他的组件,如锌指,免疫球蛋白组件,G 蛋白偶联受体和蛋白酪氨酸激酶,却不成比例地成倍增长,但是只发生在特定类别的生物体中。酵母中最大的蛋白质家族之一是 GAL4 转录因子家族(参见锌双核基因簇),但它似乎只出现在真菌类中。同样,免疫球蛋白组件只在动物中发现,并且作为蛋白质的串联重复组件特别丰富。在某些例子中,在不同界中具有相关功能的不同蛋白质家族的扩散暗示了一种随机的影响。例如,蛋白质激酶在所有的生物中都起着信号分子的作用,但在真核生物中丝氨酸/苏氨酸激酶与酪氨酸激酶是两个较大的蛋白质家族,而在细菌中它们是罕有的。相反地,组氨酸激酶在细菌中大量出现而在真核生物中并非如此。在例外的例子中蛋白质家族的扩展可能反映了进化中的功能的革新:在动物中 EGF 与免疫球蛋白组件的扩散可能反映了它们介导细胞-细胞接触的能力,这对发育中复杂的细胞相互作用有益处(见第 6 章发育,分子机制)。

22.5 蛋白质功能的综合分析

蛋白组　基因组指细胞内全部完整的遗传信息(见基因结构与作图),是细胞(或多细胞生物)管理与繁殖所要求的综合程序的储备。**蛋白组**(proteome)代表了基因组编码的蛋白质的全部集合,也即基因组的综合的功能谱。蛋白组是一个复杂的系统,虽然它只代表基因组的一部分(即编码序列)。这是由于存在不同的方式使用同一基因以产生多种产物(见第 9 章基因表达与调节)。因而蛋白组包含很多相互覆盖,结构相似的产物。蛋白组与周围环境相互作用产生**表型组**(phenome),即生物所显示的所有特征的总和的描绘。

功能基因组学　近年来，由于基因组作图与测序计划以及与之一致的了解 cDNA 序列的工作的成功(见基因结构与作图)，序列信息的数量出现了爆炸性增加。新基因序列的功能可通过与已克隆的其产物功能已了解的基因比较来试探性地假定。然而，很多新基因显示与以前了解的家族没有任何关系，功能从头分析往往是通过观察表达模式，突变的效果，与其他细胞元件的相互作用在个体基础上进行。

为利用来自测序计划的大量数据，对单个基因的深入功能分析必须被基因组范围的功能分析方法(**功能基因组学**，functional genomics)所取代，或至少被补充。这包括蛋白组中表达，突变与相互作用的系统性分析。

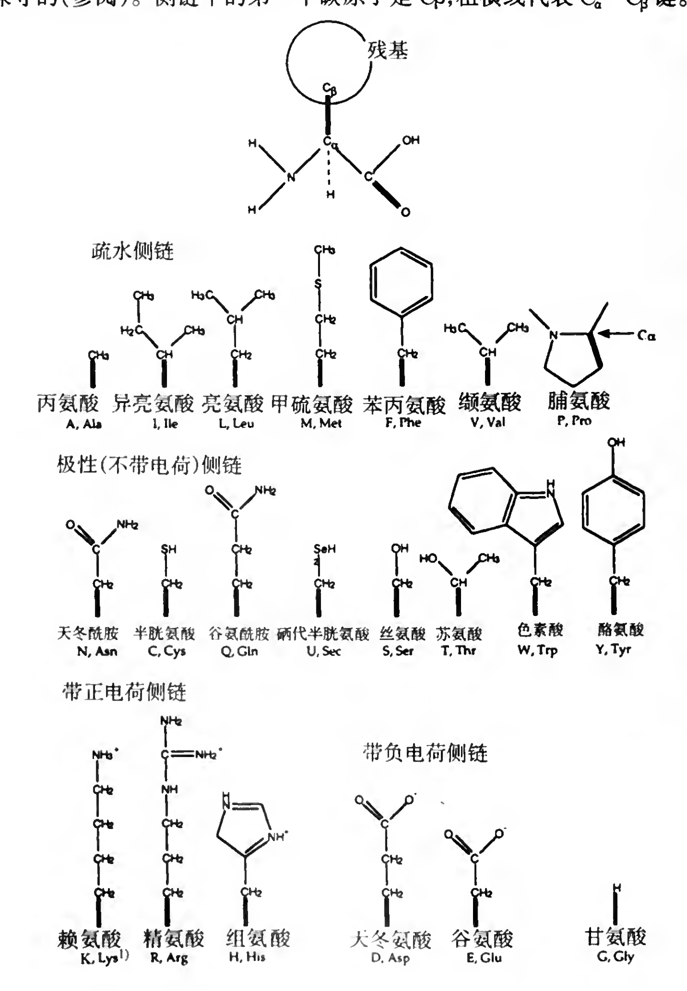
最近已发展了几种技术对基因组中所有基因的表达同时进行分析，提供细胞类别特异的基因表达图谱及细胞对环境的反映的精确描绘。**DNA 显微列阵**(DNA microarrays)与**寡核苷酸芯片**(oligonucleotide chips)使 cDNA 或寡核苷酸可以被精确地排列，杂交分析可以定性与定量地监测 RNA 的表达。**基因表达的连续分析**(serial analysis of gene expression, SAGE)技术是一以 PCR 为基础的方法，对应于特定 RNA 分子的短顺序标记被扩增和克隆，提供了标记物的一线性排列以辨别表达基因。这些技术在标记细胞反应时很有用，是高流量原位杂交等方法的补充，可提供基因表达的数据库，并帮助建立**遗传网络**(genetics networks)(见下文)。

在所有的模式生物中突变被广泛应用来确定单个基因的功能。在遗传学上了解得最清楚的生物中，所有基因的系统性破坏是可行的，并且一个国际性的实验室联合体已在全部测序的酿酒酵母中开始这项工程，用包含可被 PCR 识别的寡核苷酸盒的寻靶载体(参阅)来分别破坏 6000 个基因中的每一个。基因导向在酵母中非常有效，但在其他真核生物中就并非如此。另外，对模式脊椎动物(如小鼠)的大基因组，系统性的敲除工程要求巨大的资源投入，这种方法不太可能在较近的将来被应用。然而，有可能产生各个基因被破坏的胚胎干细胞(参阅)库。这种库可以与完整的基因组序列一起应用于提供产生缺失任何特定基因的小鼠的资源。这种方法避免了饲养与交配大量突变体的需要(参见基因捕获)。

在蛋白水平，正在发展同时监测细胞内所有蛋白的丰度与修饰状态的技术。一种方法是基于双向(2-D)电泳(文框 22.2)，另一个方法是以芯片为基础的，但以抗体代替寡核苷酸，对蛋白质的相互作用的分析是确定蛋白质功能的有力的方法。很多不同的技术可以显示蛋白质-蛋白质相互作用和蛋白质和核酸之间的相互作用，但噬菌体展示(参阅)和酵母双杂交系统(参阅)在高流量研究中特别有用，一种将酵母中所有蛋白的相互作用系统性列出的工作正在进行之中。用生化方法研究功能性的相互作用的问题之一是发现很多相互作用与生理是不相关的。这可能是由于这些产物并非是同时或在同一细胞类型中表达，即这种相互作用是外来的。那么，这些方法的综合应用——相互作用，表达模式和突变表型——是对蛋白功能彻底阐述所必须的。

文框 22.1 氨基酸

标准氨基酸 **氨基酸**(amino acid)是具有 H_2N-CHR-COOH 的普遍结构的两性(amphoteric)分子(它们可以作为酸或碱起作用)。它们在生物体内很丰富,可以以自由的二极离子(两性离子),短肽(如激素)或在蛋白质中出现。每一个氨基酸都有一个中心的四价碳原子(Cα),它具有四个配位基团,其中的三个(氨基,羧基和氢原子)是不变的,第四个基团被称为**残余基团**(residual group)或**侧链**(side chain),是可变的并是分子的物理与化学性质的决定者。存在着几百种不同类型的氨基酸,每一个都具有不同的侧链,但蛋白质是由遗传密码所对应的 21 种标准氨基酸的基本集合所形成。其中之一,硒代胱氨酸十分稀少(参见*含硒蛋白质*)。氨基酸可根据其侧链的物理化学性质归为不同类别:一些完全由碳氢基团组成,是疏水的;另一些是极性的,包含酰氨基团或羟基,可以形成氢键;其他的包含带电荷残基,可形成盐桥。下图中显示了氨基酸的普遍结构,及 21 种标准氨基酸的侧链并根据其化学性质分类。同类氨基酸的取代往往是保守的(参阅)而类间的氨基酸取代往往是非保守的(参阅)。侧链中的第一个碳原子是 Cβ,粗横线代表 C_α—C_β 键。氨基酸可用三



1) 赖氨酸单字母简称为 K,原文误写为 L。——译者注

字母或单字母简称(three letter or one letter assignment)来表示,每一氨基酸的特殊简称在图中也显示了。对有两种可能的残基也有三种说明方式:Glx,Z代表谷氨酰氨或谷氨酸,Asx,B代表天冬酰氨或天冬氨酸(这两对氨基酸在有些类型的化学分析中很难分辨)。Xaa或X表示某一残基的性质还不了解,或者在一共有顺序中是不重要的。

所有氨基酸,除了甘氨酸,由于四价Cα碳原子是手性中心存在立体异构。然而,几乎在蛋白质中出现的所有氨基酸都是L-型异构体,反映了蛋白质合成酶的特异性(这种特异性被认为是在一"冷冻进化事件"偶然发生的,相反,生物系统中很多糖是D异构体)。脯氨酸被归类为标准氨基酸之一,但是由于它包含一个二级氨基,它实际上是一种亚氨基酸。

非标准氨基酸　个别氨基酸可能会经受共翻译或翻译后化学修饰,增加蛋白质键形成能力的积累。这样的修饰可能是永久的也可能是可逆的,可能包含小的改变也可能涉及大的化学基团的添加(见正文中的讨论)。重要的是,所有这些修饰都发生在标准氨基酸插入到多肽链中后。在很多情况下,也存在特殊的氨基酸被用作插入的底物。甲硫氨酸和丝氨酸可能在插入前被修饰,分别产生*N*-甲酰甲硫氨酸和硒代半胱氨酸。在这些例子中,修饰发生在带电荷的tRNA上,使之和常见的插入标准氨基酸的tRNA不同(起始tRNA代替了插入甲硫氨酸的$tRNA^{Met}$,而$tRNA^{Sec}$代替了插入丝氨酸的$tRNA^{Ser}$)。甲酰化的甲硫氨酸与非甲酰化的甲硫氨酸在功能上是等同(封闭甲酰化反应没有效果),但硒代胱氨酸起着与丝氨酸完全不同的特殊作用,因而它被认为是一种独立的标准氨基酸,虽然它的合成是非常规的。蛋白质中插入非标准氨基酸的另一个例子是细菌细胞壁的肽聚糖中的D-甘氨酸残基。肽聚糖由糖基化的四D-甘氨酸单位由五-L-甘氨酸桥交联形成。这是一种特殊的交联,可为青霉素所阻断。

文框22.2　研究蛋白质结构的方法

研究蛋白质大体性质的方法　蛋白质可通过电泳根据其分子量或带有的电荷进行分离(参阅)。用**SDS-PAGE**(有变性剂十二烷基磺酸钠存在的凝胶电泳)可以分离不同大小的蛋白质分子。SDS带有较大的负电荷,并与蛋白质主链结合。与蛋白质结合的SDS的量与其分子大小大略成正比,众多的SDS所带的大量负电荷使蛋白质自身所带的电荷变得可忽略。这样通过胶内的孔道时,蛋白质仅仅由于大小不同而被分离。用**等电聚焦**(isoelectric focusing)可以分离带有不同电荷的蛋白质。即在pH梯度中蛋白质移动到它的等电点(pH平衡点)的方法。在**2-D电泳**(2-D electrophoresis)中这两种方法被组合使用得到高解析度的蛋白质分离。电泳方法可以用来粗分,鉴别和分离蛋白质,并使用标记蛋白来确定其分子量。用超离心方法可以更精确地确定分子量(参见沉降系数单位)。近来,质谱方法也可用于分子量的精确测定,这种方法将蛋白质溶于可快速蒸发的挥发性溶剂中喷射到质谱仪中。

蛋白质分析中抗体的应用　抗体是对识别的抗原具有高特异性的分泌的免疫球蛋白(参阅)。这种分子因而可以用来监测与纯化蛋白。这样的抗体是**多克隆**(polyclonal)(由不同的B淋巴细胞产生)往往识别抗原的多个特征或**抗原决定簇**(epitope)。通过B淋巴细胞与骨髓瘤细胞融合产生永生的淋巴细胞克隆,称为**杂交瘤**(hybridomas)能产生**单克隆抗体**(monoclonal antibodies)。杂交瘤产生一种类型的抗体并且是永生的。近来已有可能用重组DNA技术产生抗体。重组抗体可用来监测与纯化蛋白以外,还具有新的性质,如催化活性(**抗体酶**,abzymes)和作为基因治疗的工具[参见内体(intrabodies)]。

抗体可以和放射性,荧光或酶探针连接来监测电泳后固定化于固体支持物上的特定蛋白质

[**蛋白质印迹**(western blot)[1]，**免疫印迹**(immunoblot)]，或作为筛库工作的一部分[**免疫筛选**(immunoscreening)，参见*表达文库*]。同样的原理可用来原位监测蛋白，来监测蛋白质的时间与空间分布(原位免疫组化)。通过免疫共沉淀或亲和层析柱，抗体可用来纯化蛋白质。用标记的抗体可以灵敏地测定抗原的量，如酶联免疫吸附测定(ELISA)。

研究多肽链一级结构的方法 确定多肽链的一级结构最快速的途径不是通过对多肽链的直接分析，而是通过对相应的DNA的克隆与测序。然而，这并非总是可行的，cDNA序列仅仅提供了新生多肽链的序列，而功能蛋白可能经过剪切与修饰。对多肽链序列的直接分析可以利用自动**埃德曼降解**(Edman degradation)，末端氨基酸被标记并被特异地切除和识别。这个过程已被自动化，由**测序仪**(sequenators)完成，每次测定可以确定多肽链约50个残基长度的序列。较大的蛋白质可以首先用特异蛋白酶或二肽特异的化学物质(如hydroxlamine特异地切开精氨酸与甘氨酸键)水解为短的肽段进行测定。**蛋白质微量测序**(Protein microsequencing)通过气相自动化埃德曼降解接着进行高效液相色谱进行。这种方法可以对皮摩尔量的蛋白质直接测序，如聚丙烯酰氨胶洗脱的条带，从而提供一直接的途径来设计简并寡聚核苷酸，用来在一适宜的文库(参阅)分离相应的cDNA。

研究高级蛋白质结构的方法 现在还不可能从一级序列数据确定其高级结构，因此一系列生物物理的方法被用来直接监测蛋白质的三级与四级结构。

圆二色性(circular dichroism，CD)描绘了不对称分子的用左右圆偏振光吸收差谱表示的光学活性。在160到240 nm的CD光谱可以快速了解蛋白质的二级结构，因为α螺旋，β折叠和卷曲产生不同的CD谱。例如α螺旋产生具有190nm左右正的差吸收峰与210nm和220nm的负的差吸收的双峰特征的谱图。

X射线晶体衍射和核磁共振光谱技术用来研究蛋白原子结构。**X射线晶体衍射**(X-ray crystallography)使用精确定向的蛋白质晶体散射X射线到监测器上。X射线被电子散射，散射效应的幅度与原子中电子的数目成正比，光波会互相加强或削弱，其方式依赖于蛋白质中不同原子的空间排列。最终得到的影像，一系列具有不同强度的点或**反射**(reflection)，可以通过被称为**傅立叶转换**(Fourier transform)的数学函数来重建蛋白质图像。这些数据被用来确定电子密度和相位来产生电子密度图。图的解析依赖于在傅立叶合成中用到的数据的量，这也决定了最终得到的结构的解析度。X射线晶体衍射得到的最大的解析度大约为0.1～0.2nm。

中子散射(neutron scattering)可以用来提高X射线晶体衍射图像因为中子由较小的原子包括氢产生强烈的衍射。使用X射线或中子的**纤维衍射**(Fiber diffraction)可用来分析较长的纤维状分子如核苷酸和纤维状蛋白质(例胶原蛋白，角蛋白)。与晶体衍射分析能产生精确的三维图像不同，纤维衍射图式代表纤维状分子的圆柱横截面的两维平均(参见*双螺旋*)。

核磁共振光谱(nuclear magnetic resonance spectroscopy)可用来分析溶液中蛋白质的结构。其解析度与X射线晶体衍射的解析度相同，但这种技术只能应用于低分子量的分子。NMR的基础是包括氢，氮和磷的一些核(以及碳和氧的稀有的同位素)具有内源的磁性，在外加的磁场中可以通过吸收电磁能在不同的磁自旋状态之间转换[同样的技术，**电子自旋共振光谱**(electron spin resonance spectroscopy)应用于顺磁性物质即包含不配对电子的物质]。吸收可以以每一种核特异的共振频率来记录。另外共振频率受周围电子密度影响，在不同的化学环境中的原子发生**化学位移**(chemical shift)并在不同的共振频率吸收能量。这可用来鉴别不同的化学基团(甲基，芳香基等)。在磁脉冲后核磁共振衰变的方式也能提供很多信息，因为它依赖于分子的结构与空间构象。**核欧沃豪斯效应**(nuclear Overhauser effect，NOE)是磁能量在空间转移的结果并且只发生在相互作用的核相距小于0.5 nm时。**两维NOE光谱**(NOE spectroscopy，NOESY)可辨别距离较近的原子产生的对称峰，位于传统的一维谱的对角线处。**自旋-自旋偶合**(spin-spin coupling)是磁能量通过化学键或邻近核的转移。这也可以通过其他形式的二维NMR，被称为COSY或TOCSY来解析。

---

1) 蛋白质印迹是在上世纪70年代末80年代初以Southern印迹为基础发展起来的一种检测蛋白质的方法，故又称为“western(西方)”印迹，也称为免疫印迹。这种方法和Southern印迹的不同之处是：①检测的对象是蛋白质而不是核酸；②用电场的电位差将蛋白质转移到团体膜上，而不是用毛细管吸印的方法。——译者注

文框 22.3 蛋白质定位,分选和加工

胞内蛋白质运输 **蛋白质导向**(protein targeting)指指导蛋白质达到胞内特定目的地的过程(参考基因导向)。在细菌中,目的地的选择包括胞浆中,内外细胞膜,或其间的周质空间。蛋白质也可能被分泌到胞外。在真核生物中,蛋白质还可能被导向到几种胞内细胞器的任何一种,以及导向到核内。因为所有的细菌蛋白质和大部分真核蛋白质是在胞浆中合成的,被导向的蛋白质必须携带有可识别的顺序或结构使其能被转运到正确的细胞空间。这个过程被称为**蛋白质分选**(protein sorting)或**蛋白质运输**(protein trafficking)。

分选信息的本质 很多定位于特定细胞空间的蛋白质包含特异的氨基酸顺序或具有特定化学性质的残基的排列以和细胞的分选机器相互作用。这些保守残基被称为**信号序列**(signal sequences)。它们往往出现在多肽链的末端(**信号肽**,signal peptides)并且在蛋白质到达其目的地后可能被切除。具有后来被去除的信号顺序的蛋白质被称为**前蛋白**(preprotein),信号肽被称为**前序列**(presequence)。这不同于为蛋白水解所活化的被称为**蛋白原**(proprotein)的蛋白(表 22.2)(具有前序列并可被蛋白水解所活化的蛋白质被称为**前蛋白原**(preproprotein),如前胰岛素原)。当蛋白质必须穿越细胞内几个相互作用的表面时,一系列信号顺序可能按一定次序排列并可能顺次发生作用。N 端信号肽是新生肽链从核糖体上产生时出现的第一个结构,使得分选过程可能和翻译同时进行。C 端和中间的信号肽如定位核过氧化物酶体的信号肽以及在翻译后起作用的信号块在翻译后起作用。所谓的**信号块**(signal patch)指通过折叠在蛋白质表面形成的信号区域,在一级序列上并不一定连续。蛋白质在越膜之前必须保持去折叠状态或者必须先被去折叠,这个过程由分子伴侣(见正文)协助进行。信号肽一般直接与靶蛋白相互作用而信号块往往是间接发生作用,如糖基化位点这种糖残基为细胞分选机制所识别。分选信号可通过与其他蛋白的相互作用来遮蔽,这可被用作一种调节机制,例如转录因子 NF-κB 具有一核转运信号,此信号可以被抑制因子 I-κB 所遮蔽。这样调节系统通过摧毁 I-κB 可以使 NF-κB 进入核内并活化其控制的基因(见第 28 章信号传导,第 29 章转录)。

分泌途径 在真核生物中,分泌蛋白首先被定位到内质网(ER)中。这要求 N 端疏水信号肽,它可被称为**信号识别颗粒**(signal recognition particle,SRP)的胞浆核糖核蛋白复合物所识别。SRP 包含六个蛋白质和一个小的与 Alu 元件(参见此条)同源的胞浆 RNA(**7S RNA**)。当信号肽从核糖体上合成后,SRP 就与信号肽结合并使蛋白质合成停止直到复合物到达 ER。在这儿,SRP 与其受体,位于内质网膜上的二聚体**停靠蛋白**(docking protein)结合(SRP 具有三个功能结构域分别起着与信号肽结合,与受体结合,和阻断肽延长的功能)。核糖体与其自身的受体一称为 Sec61 的三聚体跨膜蛋白相缔合,蛋白合成继续进行。通过 Sec61 复合物(一些蛋白质还要求另一个元件,TRAM),多肽链进入 ER 腔,这个过程被称为**载体释放**(vectorial discharge)或**协同翻译转运**(cotranslational import)。在合成后,GTP 交换使核糖体从 SRP 上释放出来,SRP 也从其受体上释放返回到胞浆中。

当多肽链进入 ER 腔时,信号肽被十聚体信号肽酶切除。在三肽基序 Asn-Xaa-Ser/Thr[被称为**顺序子**(sequon)]有精氨酸残基出现的蛋白被寡糖单位 N 端糖基化。并非所有的顺序子都被糖基化,因而其他残基也可能与糖基化位点的决定有关,或者蛋白质的结构自身可能会抑制特定肽段的糖基化。

跨膜蛋白具有内部的疏水**转运终止顺序**(stop transfer sequence),使蛋白被锚定在内质网膜上。其余的蛋白质被包装成颗粒运输到高尔基体中,而那些应该留在 ER 腔内的蛋白质被识别并选

择性地回到 ER 中。这种蛋白质往往具有一个有共同的 KDEL 的序列**保留信号**(retention signal)。

在高尔基体中,N 连接的多糖链可以被进一步修饰并发生丝氨酸,苏氨酸和羟赖氨酸的重新 O-糖基化。在某些例子中,糖基化是蛋白质正确折叠与功能所必须的(参见*表达克隆*),而在另一些蛋白质中糖基化位点自身起着定位信号的作用。具有能被 6-磷酸甘露糖修饰的信号块的蛋白质被定位到溶酶体。另一些蛋白质被分泌,虽然其中一些具有使它们在分泌途径中停留在某一细胞器的膜上(见下文)的序列。

在细菌中,将被分泌的蛋白质以具有 N 端信号的前蛋白的形式合成,这种 N 端信号有时被称为**导肽**(leader peptides)。细菌的信号顺序较短(≤25 残基)包括一疏水中心核,可以采取 α 螺旋结构,两侧区域具有几个带电荷的残基。一些分子伴侣蛋白可以在前导肽从核糖体中出现时与新生肽链结合,以防止错折叠。其中之一,Sec B 在蛋白质转运中起着主要作用,因为它与分泌系统的另一成分 Sec A 结合,Sec A 是与易位机制有关的分子伴侣。Sec A 通过以 ATP 水解依赖方式将底物蛋白送到易位装置(包括跨膜蛋白 Sec E 和 Sec Y)中来介导易位。这时导肽被称为**导肽酶**(leader peptidase)切除。在成熟蛋白质中,在剪切位点邻近(-1 位置)和下游再隔一个残基的位点(-3 位置)往往会是具有小的残余基团的氨基酸。这个现象被称为 **-1 与 -3 规则**(-1 and -3 rule)。

定位蛋白质到膜上 在真核生物中,进入分泌途径的很多蛋白质是将被滞留在特定的膜上的。转运终止顺序使蛋白质可被保留到 ER 膜上,通过组合不同的类似顺序(**膜锚定信号**,membrane anchor signals)与特定的保留信号,蛋白质也可被定位到高尔基体与溶酶体膜上。任何缺少保留信号的膜结合蛋白将会到达细胞膜,但这并不是蛋白质到达这个目的地的唯一途径。在高尔基体中的带有 **GPI 顺序**的蛋白将与 GPI(糖基磷脂酰肌醇)锚连接,它将使蛋白质到达内膜表面。被脂肪酰化(如豆蔻酰化)的胞浆蛋白质也被定位到细胞膜上。

转运到核和胞内细胞器的蛋白质 小的蛋白质可通过核孔持续扩散,但定位到核的较大的蛋白(>50kDa)就必须被转运,转运过程包括**核定位顺序**(nuclear localization sequence, **NLS**)的识别。NLSs 往往较短,富含脯氨酸与碱性残基,而缺少疏水残基。一些 NLSs 是连续的,例如 SV40T 抗原 NLS(**Pro-Pro-Lys-Lys-Lys-Arg-Lys**),而另一些是由两部分组成的,由一段极性残基将两段富含脯氨酸/碱性区域分开(例如 p53 NLS 为 **Lys-Arg**-Ala-Leu-**Pro**-Asn-Asn-Thr-Ser-Ser-Ser-**Pro**-Gln-**Pro-Lys-Lys-Lys**)。二聚体胞浆蛋白质,**输入蛋白**(importin),是核输入的附着阶段所必需的。输入蛋白 α 与底物的 NLS 顺序结合,输入蛋白 β 识别和结合核孔成分,核孔蛋白(necleoporin)。一种被称为 Ran 的 GTP 酶引发输入蛋白解聚。在核转运过程中可能包括一系列这样的反应使附着在核孔蛋白上的底物进入核孔。

即将被转运到线粒体或叶绿体的蛋白可以被定位到不同的空间内,内膜的外边,膜间空间,内部介质/气孔。在叶绿体中可定位到类囊体[1]膜或腔。转运进细胞器的蛋白质常常包含一 N 端**转运顺序**(transit sequence)。这是一 15~70 残基顺序,富含极性与碱性残基(叶绿体转运顺序比线粒体转运顺序极性残基含量要高),被细胞器表面的受体所识别。在线粒体的介质或叶绿体的气孔处,转运顺序被剪切,使蛋白质定位在这些地方。蛋白质在内膜和外膜处与转运机器相互作用,同时穿过内膜和外膜被转运。定位到其他的空间要求进一步的信号。细胞器定位信号以顺序方式排列,所以 N 端转运顺序被切除可能活化不同的信号。切除转运顺序的多肽酶位于介质/气孔处,所以即使细胞器蛋白的最终目的地不在此处,它们也必须暂时地进入介质或气孔。插入膜的蛋白

1) 原文误写为"thykaloid",应为"thylaloid"(类囊体)。——译者注

质往往包含疏水的终止转运顺序。

蛋白质降解途径　　蛋白质的稳定性相差很大，其半寿期从几分钟（如很多调节蛋白质）到几个星期或更长（例如很多结构和储存蛋白：胶原蛋白，血红蛋白）。在细胞中存在着几种不同的降解途径，其活性根据营养状态和可能涉及蛋白质的修饰的其他细胞的信号而不同。很多不稳定特征的蛋白质都包含 **PEST 位点**（PEST sites，即富含脯氨酸，谷氨酸，丝氨酸和苏氨酸）。另一些蛋白质可能通过翻译后修饰被定位到降解途径，或者包括特定残基的化学修饰（如磷酸化）或与小的蛋白**泛素**（ubiquitin）连接。蛋白质的 N 端残基及其修饰状态可能通过作为泛素形成潜在的靶在蛋白质代谢的调节中起重要的作用。这被称为 **N 端法则**（N-end rule）。

溶菌途径是泛素依赖途径，往往包括相对稳定的蛋白质。蛋白质通过颗粒出芽而被内化，这个过程被称为微自体吞噬（microautophagy）。在饥饿状态，某些胞浆蛋白可能可以通过信号肽的识别直接进入溶酶体。

泛素依赖降解，包括小蛋白泛素的活化是胞质蛋白质的主要降解途径，泛素可以由载体转运到其底物，并在一个称为**蛋白酶体**（proteasome）的大的多蛋白复合物中将底物定为降解的靶。泛素通过与 E1 蛋白质连接而被活化，然后被转移到蛋白质 E2 中，接着转移到其靶蛋白上。靶蛋白由与蛋白质 E3 结合来辨别（**泛素连接酶**，ubiquitin ligase）。底物的特异性由 E2 和 E3 共同控制，存在几种同工形式可能定位不同类型的蛋白质。E2 转运泛素到靶蛋白的亮氨酸的 ε-氨基上。然后泛素自身也可成为泛素化的对象，连接位点为 Lys46，产生多泛素化的底物，是蛋白酶体降解的信号。

（徐　沁译）

第23章
蛋白质合成

基本概念和定义

- 蛋白质合成是基因表达中信使 RNA(mRNA)所携带的遗传信息被**翻译**为**多肽链**的阶段。
- 蛋白质合成装置的组分有:mRNA,包含被翻译密码的模板,**核糖体**,大的核糖核蛋白颗粒(是蛋白质合成的位点),转运 RNA(tRNA),种类较多的**衔接分子**(adaptor molecular)携带氨基酸到核糖体上以推进翻译过程,以及与核糖体暂时缔合的**附属因子**,为对于核糖体装配与解离及翻译延伸时的活性所必需。
- 与其他的多聚反应类似,蛋白质合成具有**起始**(initiation),**延伸**(elongation)与**终止**(termination)等阶段,其中每一阶段都可被调节蛋白质合成的机制在真核生物与原核生物中是相同的,只在组分的性质与它们装配的次序上有微细的差异。但在蛋白质合成发生的前后却有较大的差别。在细菌中,转录与蛋白质合成在胞浆中同时发生(使不同基因表达水平之间的交互调节可能发生)并且 mRNA 可能为多顺反子。在真核生物中,转录被限制在核中,mRNA 被转运到胞浆中进行翻译。新生 mRNA 在转运前被进一步加工,并且常常是单顺反子(见第 27 章 RNA 加工)。
- 在合成后,多肽链经历进一步的加工过程以具备生物活性。它必须正确**折叠**,这个过程往往需要分子伴侣(参阅)的帮助。它可能被剪切,特定的残基可能被化学修饰或与小分子连接。这样的修饰往往与蛋白质定位到细胞内的特定空间或分泌有关。蛋白质的完整的生物活性可能还要求与其他蛋白质或与非多肽辅因子非共价缔合。对这些过程的讨论,见第 22 章蛋白质:结构,功能与进化。

23.1 蛋白质合成的组分

信使 RNA　　**信使 RNA**(mRNA)是蛋白质合成的模板。它具有两个必需特征:**开放的阅读框**(一段可翻译密码序列)和一个**核糖体结合位点**(ribosome binding site,核糖体小亚基结合核糖体的其他部分装配的位置)。在这些位点的组织上,原核生物与真核生物有重要的差别,这也关系到一典型的 mRNA 的寿命的其他方面。这些差别与其后果在下面作一总结。

(1) 在细菌中,转录与翻译在相同的细胞空间同时发生,而在真核生物,转录被限制在核内,RNA 被转运到胞浆中进行翻译。

(2) 细菌 mRNA 具有有限的半寿期(最稳定的转录产物为几分钟)。一些真核 mRNAs 也不稳定,但大部分在几小时甚至几天内是稳定的(例如在卵中)。

(3) 细菌转录产物被直接用于翻译,而真核转录产物在翻译前被进一步加工和修饰。加工反应包括内含子的剪切和 3′端的多聚腺苷酸化,两者都可能调节翻译的效率,或者直接影响或者通过调节 mRNA 的稳定性。进一步的修饰反应是 5′端 7-甲基鸟苷酸帽结构的合成,在核糖体结合中起直接的作用。一些转录产物还可被编辑(参见 *RNA 编辑*)。

(4) 细菌中与核糖体的结合依赖于 mRNA 中的一保守基序,它与核糖体 16S rRNA 的部分互补。结合可发生在此顺序出现的任何地方,包括顺序中间,因此细菌转录产物可能是多顺反子的。真核生物中核糖体通过一识别被修饰的 5′帽结构的蛋白质而锚定在 mRNA 上,这样结合不可能发生在 mRNA 的中间,所以真核 mRNAs 几乎都是单顺反子的(然而某些 RNA 病毒可以克服这种限制以多顺反子形式表达它们自身基因组中的基因;参见*内部核糖体进入位点*)。典型的细菌与真核 mRNAs 的结构在图 8.2 中显示(见第 8 章基因)。

这些差别的广泛后果反映在原核生物与真核生物使用的调节策略的不同。细菌 mRNAs 快速地被转录,翻译和降解,并且这三个过程的紧密关系使得基因表达的不同水平之间的大量交互调节的发生(参见*衰减作用*,*反向调节*)。反之,在真核生物中,转录与翻译过程之间有明显的延迟,mRNA 首先被加工然后从核中转运出来。这两个过程都是调节的潜在作用过程。

核糖体　蛋白质合成发生在**核糖体**(ribosomes)上,它是大的含量丰富的核糖核蛋白复合物。它们在胞浆中以游离的形式存在,在真核生物中,还发现它们缔合在粗面内质网膜上。核糖体可以单独产生功能(**单核糖体**,monosomes),但是更常见的是它们成簇同时翻译同一 mRNA(**多核糖体**,polysomes),多核糖体可以从细胞中抽提而且常常被用来纯化 mRNA(参见 poly(A)$^+$RNA)。

所有的核糖体由两种不同大小的亚基组成,**大亚基与小亚基**(large and small subunit)。每一亚基由几种**核糖体 RNAs**(ribosomal RNAs, rRNAs)和多种**核糖体蛋白质**(ribosomal proteins, r-proteins)组成。它们的相对大小常常用沉降系数单位来表示(文框 23.1)。在大肠杆菌中,70S 核糖体由小的 30S 亚基和大的 50S 亚基组成。小亚基包含 21 种不同的蛋白质(被称为 S1～S21)和 16S rRNA。大亚基由 34 种蛋白质(被命名为 L1～L34)和 23S 及 5S rRNAs 组成。有些蛋白质在两种亚基中是相同的(如 L16=S20)。真核核糖体较大(80S)并且包含更多组分。小亚基(40S)包含 33 种蛋白质和 18S rRNA,而大亚基(60S)包含 50 种蛋白质和 3 种 rRNA,28S,5.8S 和 5S。5.8S rRNA 与细菌 23S rRNA 的 5′部分同源并且与等同的真核 28S rRNA 形成互补配对。古核糖体类似于细菌核糖体,但有些包含与真核相同的特别亚基。

核糖体的空间组织是复杂的。rRNA 组成总分子量的 60%～65%,对结构完整与功能都是必需的,它们通过分子内或分子间碱基配对采取复杂的三级与四级构象。利用交联与足迹法,以及中子散射和衍射进行的结构分析详尽地研究了大肠杆菌核糖体中蛋白质与 RNA 和蛋白质之间相互作用的方式。虽然不同种属中 rRNAs 的一级核苷酸顺序有显著的不同,大部分二级结构似乎是保守的,表明所有的核糖体具有共同的组织方式。

核糖体的几个结构域在蛋白质合成过程中具有特别的重要性。小亚基包含 mRNA

的结合位点和 tRNA 的两个主要结合位点。**A 位点**（**氨基乙酰 tRNA 位点**，amino acyl-tRNA site）在延长过程中与进入的带电荷 tRNA 结合，而 **P 位点**（**肽酰 tRNA 位点**，peptidyl-tRNA site）与携带新生多肽链的 tRNA 结合。细菌核糖体具有第三个 **E 位点**（**离开位点**，exit site），空载的 tRNAs 从此位点被排出。大亚基具有**肽基转移酶结构域**，提供肽键形成的催化活性，GTP 酶结构域，其活性是核糖体在 mRNA 易位所必需的。这些位点在蛋白质合成中的作用将在下面详细地讨论。

转运 RNA　　在遗传密码被了解以前，弗朗西斯·克里克（Francis Crick）提出**转接器假说**（adaptor hypothesis）来解释 mRNA 中的核苷酸顺序如何被翻译为蛋白质。这个模型预测了既能识别信息的核苷酸序列又能识别适宜的氨基酸的转接器分子的存在，它将两者都携带到核糖体上。

这种转接器分子是**转运 RNA**（tRNA）。tRNAs 是同源性相对较高的 RNA 分子，长度在 75～100 残基之间，在它们产生后将被进一步地加工（见第 27 章 RNA 加工）。它们具有特殊的二级与三级结构（文框 23.2）。最重要的有**受体臂**（acceptor stem，氨基酸结合位点）和**反密码子环**[anticodon loop，携带有与 mRNA 上的密码子配对的互补碱基的三核苷酸**反密码子**（anticodon）]。细菌细胞包含 35 种不同的 tRNAs，真核生物中超过 50 种。这个数目低于遗传密码中可能的密码子数目，但大于密码子所代表的氨基酸数目。这表明了个别 tRNAs 可以识别一个以上的密码子（反映了*摆动配对*，参阅），但不同的 tRNAs 有可能装载同样的氨基酸（即*同工* tRNA，参阅）。

tRNAs 被一种称为**氨酰 tRNA 合成酶**（amino acyl tRNA synthetases）的蛋白所**装载**（与相对应的氨基酸连接）。对应每一种氨基酸有一种酶，因此每一种合成酶识别所有的同性质的同工 tRNA。装载机制与其调节在别处详细讨论（见第 11 章遗传密码）。

23.2　蛋白质合成的机制

总述　　蛋白质合成的起始阶段包括蛋白质合成元件的聚合，核糖体定位到翻译起始位点，以及多肽链的第一个氨基酸的放置（图 23.1）。当第一个多肽键形成时起始结束。在真核生物与原核生物中，核糖体的小亚基与 mRNA 结合早于大亚基的结合，其结合要求一系列的**起始因子**（initiation factors，IFs），它们在核糖体大亚基被征集时释放。小亚基可识别 mRNA **核糖体结合位点**（ribosome binding site）。第一个氨基酸与一特异的起始 **tRNA**（initiator tRNA）连接，它具有进入小核糖体亚基的部分 P 位点的特性。起始密码子常常是 AUG，它标志着开放读码框的开始并且设定了多肽链其他部分的读码框。所有其他的**延伸 tRNAs**（elongator tRNAs）在延伸阶段进入完全装配好的核糖体的完整 A 位点。延伸包括三个反应的循环：征集，转肽和易位。每一步都要求**延伸因子**（elongator factors，EFs）的作用与 GTP 的水解来提供能量。三个反应的净效应是将胞浆中的氨基酸转移到新生肽链上并使核糖体以三个核苷酸残基为单位沿着 mRNA 移动。当核糖体遇到一终止密码子时蛋白质合成停止，核糖体上附着的因子被释放，翻译装置从而解散。

细菌中的翻译起始——详细的机制　　在真细菌中，30S 核糖体亚基只在与起始因子 IF3 缔合后才能结合 mRNA，IF3 可以稳定小亚基的结构并防止与大亚基的成熟前相互作用。

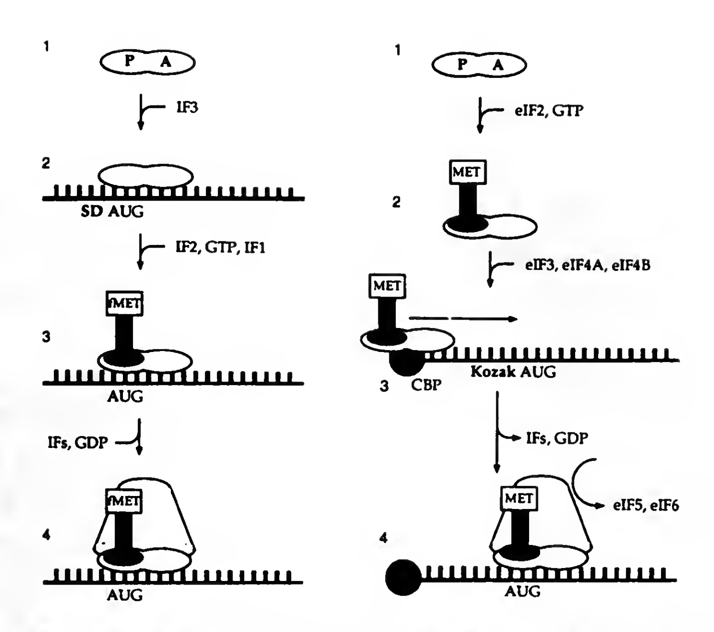


图 23.1 细菌与真核生物中的蛋白质合成的起始。在细菌中(左)小亚基(1)与 Shine-Dalgano (SD)顺序结合(2)并征集起始 tRNA (3) ,大亚基从而与之结合(4)。在真核生物中(右),小亚基(1)与 eIF-2-GTP 及起始 tRNA 结合形成前起始复合物(2),此复合物可识别 mRNA 的 5′帽子上的帽结合蛋白(CBP)从而结合 mRNA 并沿途搜寻合适位置上的起始密码子为止(3),然后大亚基与之结合(4)。为了显示清楚,图中小亚基只覆盖 6 个碱基,而实际上它覆盖 30~40 碱基,这样它能同时与细菌 SD 顺序(或真核生物的 Kozak 顺序)及起始密码子相互作用。核糖体的 A 位点和 P 位点在图中显示。

虽然 IF3 控制着小亚基与 mRNA 结合的能力,识别的特异性仍是由核糖体自身提供的。mRNA 上的核糖体结合位点是 **Shine-Dalgarno 顺序**(保守顺序 UAAGGAGG),它与核糖体 16S rRNA 中的互补序列形成碱基对。起始密码子(一般为 AUG 但有时是 GUG 或 UUG)处于 Shine-Dalgarno 顺序下游约 10nt 处,核糖体的正确定位是起始密码子处于核糖体亚基的 P 位点。

起始密码子对应于氨基酸的甲硫氨酸。在大肠杆菌中有两种携带甲硫氨酸的 tRNAs。其中之一($tRNA_f^{Met}$)特异地识别起始密码子,而另一个($tRNA_m^{Met}$)识别中间的 AUG 密码子(内部的 GUG 和 UUG 密码子分别被 $tRNA^{Val}$和 $tRNA^{Leu}$识别)。一旦起始 tRNAs 被装载,甲硫氨酸的氨基就会被甲酰化修饰。这一修饰的信号可能是双链受体臂末端的不配对残基,相应的残基在延伸 tRNA 中是配对的(甲硫氨酰的 $tRNAs_m^{Met}$不能被甲酰化)。这一信号还可防止在延伸过程中起始 tRNA 进入核糖体的 A 位点,而且此修饰封闭了氨基基团从而防止 N-甲酰甲硫氨酸与上游残基形成肽键。在起始过程中,起始 tRNA 被装载到小核糖体亚基的 P 位点。P 位点提供了特殊的分子环境,只有起始 tRNA可以进入,而延伸 tRNA 不能进入。起始密码子的特异性与反密码子环中的三个 G:C碱基对有关,这在延伸 tRNA 中是不存在的。N 甲酰化对起始特异性并非必要——包含正常甲硫氨酸的起始 tRNAs 与包含甲酰甲硫氨酸的修饰起始 tRNAs 同样有效。反密码子第三位有一定程度的摇摆(参阅)允许 AUG,GUG 和 UUG 作为起始密码子,但是 $tRNA_f^{Met}$识别这些密码子的效率较低。因此可以通过选择起始密码子来对蛋白质合成的

效率建立组成型的控制。

在 GTP 存在下，起始 tRNA 由 IF-2 携带到 P 位点。这时 IF-1 与完整的起始复合物结合以稳定它。当 50S 大亚基被装配时，所有的起始因子被释放。这一过程要求 GTP 的水解来提供能量。

真核的起始——详细的机制　真核与细菌的机制是相同的，其主要的区别在于各种成分装配的次序，真核起始因子数目较多，真核翻译的起始和延伸都使用正常的甲硫氨酸。这样起始与延伸的甲硫氨酰化 tRNAs（$tRNA_f^{Met}$和 $tRNA_m^{Met}$）携带同样的氨基酸，只在 tRNA 自身的结构上有区别。

在真核生物中，小核糖体亚基在没有和起始 tRNA 缔合时不能与 mRNA 结合，由 $tRNA_f^{Met}$，eIF-2 和 GTP 组成的三元复合物结合小亚基形成**前起始复合物**（preinitiation complex），它可与 mRNA 结合。核糖体结合位点由起始密码子所决定，虽然附近的顺序也可能十分重要，被称为 **Kozak 共有序列**（ACCAUGG）。开始，前起始复合物结合 5′7meG帽结构并沿着 mRNA 搜寻第一个起始密码子。这个过程需要几种起始因子，包括帽结合蛋白质（eIF-4F），eIF-3，eIF-4A，eIF-4B。帽结合蛋白促进 40S 亚基的起始结合。eIF-4A 与 4B 共同作用去除 5′非翻译区域（UTR）的二级结构，然后它们和 eIF-3 帮助 40S 亚基结合以形成**起始复合物**（initiation complex）。大亚基与另一个起始因子 eIF-6 相缔合大亚基与起始复合物的结合包括 eIF-2 和 eIF-3 从前起始复合物上解离及 eIF-6 从大亚基上解离，这个过程由 eIF-5 介导。起始因子 eIF-2 和 eIF-6 都是**抗缔合因子**（antiassociation factors）：它们防止核糖体亚基在胞浆中相互作用。当大亚基结合时其他的起始因子从核糖体上解离，此过程要求 GTP 水解。

延伸循环　延伸循环在细菌与真核生物中基本相同，以三阶段发生——征集，转肽作用和易位（图 23.2）。

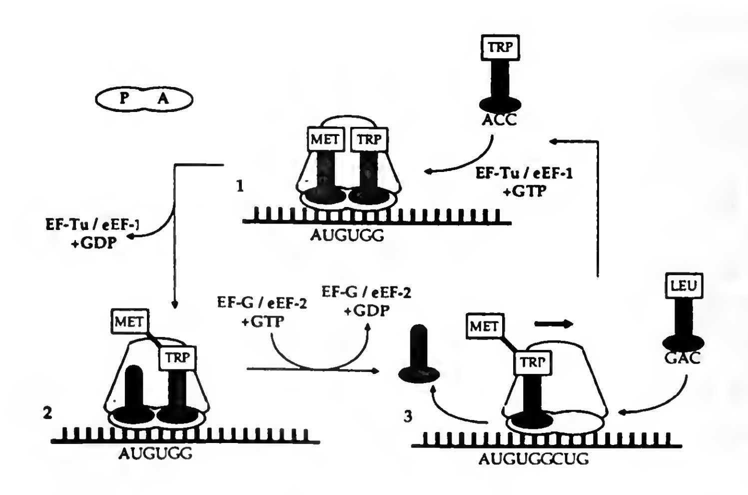


图 23.2　蛋白质合成的延伸循环。（1）征集：负载 tRNA 进入核糖体 A 位点。（2）转肽作用：连接氨酰残基与 P 位点 tRNA 的肽键被转移到填充 A 位点的氨酰残基上。（3）易位：空载的 tRNA 从 P 位点排除，核糖体向 mRNA 下游移动三个碱基，将 A 位点 tRNA 移动到 P 位点，下一个密码子进入 A 位点，为下一负载 tRNA 的征集作准备。图中标出了核糖体的 A 位点与 P 位点。

在起始后，起始氨基酰-tRNA 填充了核糖体 P 位点，下一个密码子排列在 A 位点。**第一个延伸氨基酸-tRNA** 进入 A 位点，其反密码子与密码子配对，这种识别由**延伸因子**承担(在细菌中是 EF-Tu，在真核生物中是 eEF-1，它们都与 GTP 缔合)。

这时 P 位点的连接起始 tRNA 与甲硫氨酸的肽键断裂，甲硫氨酸与充填 A 位点的氨基酸之间形成肽键。起始 tRNA 被**释放**(dischraged)。**转肽作用**(transpeptidation)由核糖体大亚基的肽基转移酶活力介导。与 EF-Tu(或 eEF-1)缔合的 GTP 被水解且延伸因子被排除。EF-Tu/GDP 是没有活力的，此核苷酸被第二种延伸因子 EF-Ts 所取代，而 EF-Ts 又可被 GTP 取代。这样 EF-Tu 可以被重复使用。在真核生物中似乎没有 EF-Ts 的对应物，不过 eEF-1 是几种多肽链的积聚，其中之一可能具有 EF-Tu 的功能。

下一阶段是**易位**(translocation)。填充在 P 位点的起始 tRNA 被排除(在细菌中被转移到另一个被称为 E 位点核糖体结构域，而在真核生物中 tRNA 从核糖体上解离下来)。甲硫氨酰残基通过肽键与填充 A 位点的氨基酸连接，这时在 A 位点的 tRNA 被称为**肽酰- tRNA**(peptidyl tRNA)。然后核糖体沿着 mRNA 移动 3 个碱基，肽酰 tRNA 仍然与其密码子缔合。肽酰 tRNA 被转移到 P 位点，mRNA 沿核糖体移动使下一个密码子处于 A 位点。易位步骤的几何学依赖于 mRNA 与 tRNA 的相互作用：具有四个反密码子碱基的突变型 tRNA 引起四个碱基的易位，导致*移码*或*移码抑制*(参阅)。易位要求另一种延伸因子(细菌中的 EF-G，真核生物中的 eEF-2)，同样与 GTP 结合并要求 GTP 的水解。在大肠杆菌和某些其他的细菌，易位还要求几种游离的 rRNA，4.5S rRNA，它们的作用仍不清楚。

蛋白质合成的终止　当核糖体遇到三种终止密码子(参阅)之一时，蛋白质合成终止。它们往往没有相应的 tRNA，使翻译停止(参考*硒代半胱氨酸插入，无义抑制基因*)。翻译完成的多肽链被释放，tRNA，mRNA 和核糖体亚基解离，在以后的翻译过程中被循环使用(图 23.3)。终止要求一系列蛋白质**释放因子**(release factors, **RFs**)，它们与 A 位点结合并识别无义密码子。它们促进了多肽链从最后一个肽酰 tRNA 上释放出来，空载的

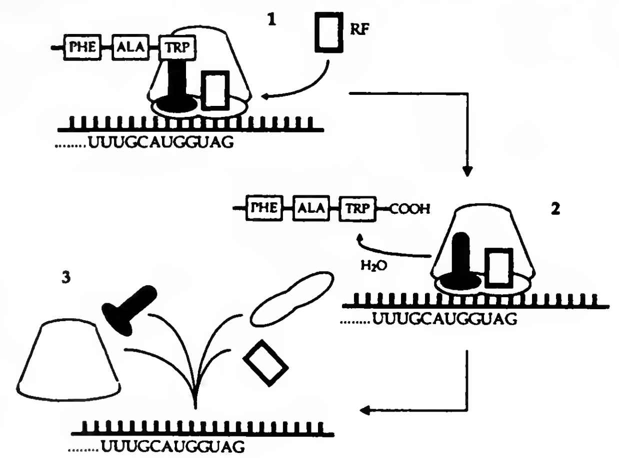


图 23.3　蛋白质合成的终止。当核糖体遇到一终止密码子时，释放因子与 A 位点结合。它将新生肽链从 P 位点的肽酰- tRNA 释放并使核糖体解离，使蛋白质合成的成分可以被循环利用。

tRNA从核糖体上离去,核糖体解离。在大肠杆菌中有两种不同的释放因子识别不同的两对终止密码子。RF1识别UAA和UAG,而RF2识别UAA和UGA。它们都需要第三个因子,RF3的帮助。在真核生物中,单个GTP依赖因子(eRF)识别所有的终止密码子,但效率不同。有时会发生通读,特别在弱终止子时,这反映了释放因子与tRNA之间的竞争。终止密码子的错读是翻译调节的重要方面(在下面讨论)并可能受mRNA的二级结构的影响。

23.3 蛋白质合成的调节

蛋白质合成的组成型控制 蛋白质合成的效率受组成性控制,但也可能在整体或在个别转录本的水平上被调节。蛋白质合成的组成型水平依赖于mRNA结构,还反映了mRNA稳定性,序列和核糖体结合位点附近序列,二级结构的出现,起始密码子的选择和整个开放读码框的密码子偏向等因素的影响。

二级结构和终止密码子的选择产生被通称为**程序性错读**(programmed misreading)的调节现象,使密码子顺序的正常翻译被抑制。程序性错读的例子包括通读、硒代半胱氨酸插入和移码。**通读**发生在弱终止密码子上(即终止密码子处释放因子被限制),在释放因子与具有可耐受的 反密码子顺序的tRNA之间存在着竞争。硒代半胱氨酸插入发生在终止密码子UGA处,并要求一定的二级结构,即硒代半胱氨酸插入序列(参阅)。当核糖体在二级结构或稀有密码子处暂停,向前或向后滑动时一个核苷酸或在一不同的读码框持续翻译会发生**移码**(frameshifting)或**再编码**(recoding)。这可以用来在翻译中途改变读码框(例如在反义病毒*pol*基因)来导致早期平截(trucation)(例如MS2 *lysis*基因)或防止平截来延伸一基因产物(例如在大肠杆菌中*dnaX*基因)。

在细菌中,操纵子环境可产生翻译调节的新的形式,下游基因的翻译依赖于多顺反子中上游基因的翻译。个别开放阅读框的空隙十分重要。连续开放阅读框之间的空隙大于大约30核苷酸,翻译以分立的起始事件开始,每一位点是独立的。如果存在较短的空隙,核糖体可以通过连接开放阅读框重新开始,即不发生从模板上第一次解离。起始信号与正常的SD-顺序不同,核糖体亚基以典型的方式和较低的效率自发装配。封闭或阻断上游基因翻译的突变也会影响下游基因的翻译,这被称为**极性突变**(polar mutation)(例如并参见*lac*操纵子)。

真核mRNA的翻译常常是帽依赖的,并通过搜寻第一个起始密码子而发生。小RNA病毒家族提供了一个例外,其基因组包含**内部核糖体进入位点**(internal ribosome entry site, **IERS**),即形成一定二级结构的基序允许蛋白质合成的内部起始。内部起始依赖于帽结合蛋白eIF-4F,当此蛋白质失活时会发生小RNA病毒内部起始(通过抑制eIF-4F阻断宿主蛋白质合成作为感染策略的一部分)。然而,IRES功能不需要病毒蛋白,即起始依赖于其他的寄主起始因子。内部起始可能发生在特定的细胞转录本,包括果蝇触角足和小鼠*fgf2*。IRES基序的丰度和它们在基因表达控制中的显著性还不了解,它们被用功能术语定义,没有显示保守的结构特征。

蛋白质合成的整体调节 蛋白质合成的"基本装置"的成分(起始,延伸和终止因子,以及核糖体的组分)可以被调节,从而对蛋白质合成进行整体的调节。例如,几种真核病毒

通过磷酸化起始因子 eIF-2 关闭寄主蛋白质的合成，当要求整体合成抑制时这个策略也可用于细胞自身(例如当细胞经受热休克时)。在酿酒酵母中 eIF-2 的磷酸化允许搜寻的遗漏，核糖体可以跳过弱的起始子而与强起始子结合。这改变了某些基因的抑制如 *GCN4*，它们在 mRNA 的帽和确定起始密码子之间有多个无效的 AUG 密码子。当 eIF-2 去磷酸化，核糖体试图在第一个 AUG 处起始蛋白质合成，而由稍后的读码框内的终止密码子所终止。

蛋白质合成的个别结构域调节 特异 mRNAs 可以被单独地调节。详细了解的例子包括哺乳类铁蛋白 mRNA 和大肠杆菌中的核糖体蛋白质 L5 mRNA。

在铁蛋白系统中，当胞内铁的浓度较低时铁蛋白 mRNA 的翻译被阻断。这种抑制依赖于转录本中的**铁效应元件**(iron-response elements, IREs)并由 IRE 结合蛋白介导，这种蛋白在铁存在时失活。铁蛋白的 IRE 出现在转录体的 5′非翻译区，IRE-BP 通过阻止核糖体的搜寻来抑制蛋白质的合成。在铁蛋白 mRNA 中，IREs 不与富含 AU 的 mRNA 去稳定位点缔合，并不影响 mRNA 的转换，如在转铁蛋白受体 mRNA 中的情况(见第 27 章 RNA 加工)。

L5 系统代表了核糖体组件合成的反馈控制。核糖体蛋白质 S8 与 16S rRNA 残基 588～651 形成的茎环结构结合。L5 核糖体蛋白质 mRNA 的 5′非翻译区与编码区域的前 30 个碱基形成同样的结构。蛋白质 S8 与 L5 mRNA 结合(虽然结合效率低于 16S rRNA)并抑制蛋白质的合成。多余的 16SrRNA 可结合所有的 S8 蛋白质，从而使 L5 的翻译去抑制，而如果 16SrRNA 的量有限(因而不需要 L5 的翻译)，过剩的 S8 蛋白质阻止 L5 的翻译。

特异翻译调节的最后一个例子由反义 RNA 提供(参阅)。这主要在原核系统中发现(例如在细菌噬菌体 λ 的 Q 蛋白质合成的控制，质粒复制基因的控制，以及不同的转座子的转座酶的合成的控制，见病毒，质粒，可移动遗传因子各章)，但在真核生物中也有发现(例如在鸡肢芽 Fgf-2 合成的控制)。最近，翻译的反义抑制被用作阻断基因功能的策略而不必进行体外或定点突变，这样无需破坏基因就可以确定零表型效应(参见转基因生物，基因治疗，共抑制)。

文框 23.1 大分子质量测定

相对分子量 分子的质量常常用**相对分子量**(relative molecular mass, M_r)来表示，指相对于碳原子的十二分之一[**一原子质量单位**(atomic mass unit)或**一道尔顿**(dalton)]的质量。因而蛋白质或核酸等大分子的质量常常用**千道尔顿**(kilodalton, kD, kDa)来表示，这等于 1000 个原子质量单位。理论上分子的 Mr 可由每一原子的加和而得到，或通过实验方法与已知的蛋白标准比较大小得到(如通过凝胶电泳)。

核酸 核酸往往比蛋白质大，常常以它们包含的核苷酸单位的数目来衡量。单链核苷酸的数目用**碱基**(bases, b)数或**核苷酸**(nucleotides, nt)数来表示，较大的分子用**千碱基**(kilobases, kb)表示。双链分子的大小用**碱基对**(bases pairs, bp)和**千碱基对**(kilobases pairs, kbp)表示，对大的 DNA 分子

如染色体，用**兆碱基对**（mega bases pairs, Mbp）表示。较之蛋白质核酸的单体分布更为平均，质量更为均一。因而从碱基数或碱基对数得到的相对分子量具有一定的精确性。

沉降系数单位　大分子与细胞成分的质可从由超离心得到的**沉降系数**（sedimentation coefficient）来估算。它以秒（s）为单位，反映了在一单位场中沉降速率。大分子的沉降系数在 1×10^{13} 到 2×10^{15} 的范围内。1×10^{13} 秒被定义为一个**沉降系数单位**（Svedberg unit, S）。沉降系数单位常常用来表示核糖核蛋白颗粒和 RNA 分子的质量，如核糖体 RNA 和核糖体亚基（另见**剪接体，信号识别颗粒**）。

换算　M_r 可根据下式由沉降系数决定：

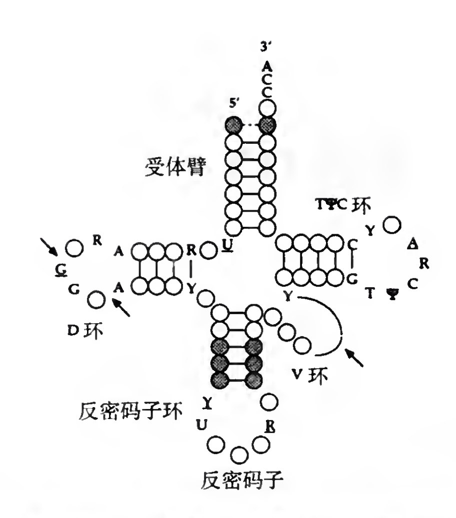
$$M_r=\frac{RTs}{D\ (1-v\rho)}$$

R 是气体常数，T 是温度，s 是沉降系数，D 是扩散常数，v 是部分特异体积，ρ 是溶剂密度。作为一种根据实际经验而得到的作法（a rule of thumb），DNA 的一个碱基对的平均分子量被认为是 650 Da，RNA 的一个碱基为 335 Da。这些是核苷酸单元的平均分子量再每核苷酸加上一个磷酸基团。这可以用来从相对分子量或沉降系数单位估算碱基或碱基对长度，或反之。

文框 23.2　tRNA 的结构与功能（又见核苷酸结构）

一级结构　所有的 tRNAs 有同样的长度，75～100 nt，并具有 10％的修饰或超修饰碱基。tRNA 中大约 20％的碱基是**不变的**（invariant）或**半不变的**（semi-invariant）（即碱基的类型，嘌呤或嘧啶是保守的），并在三级结构的形成和与翻译装置的其他元件的相互作用中起重要作用。超过一半碱基是可变的但形成分子内的配对碱基对。对这些碱基来说，形成碱基对的能力而不是碱基自身的特异性对 tRNA 的结构和功能更为重要。在不配对的环区发现其他的可变碱基，包括反密码子环。一些可变碱基使每个 tRNA 为特异的**氨酰 tRNA 合成酶**（参阅）提供了识别信号。在某些 tRNAs中，反义密码子环在特异性方面起重要的作用。但在另一些中特异性仅仅包括外部残基，其取代会改变 tRNA 的"身份"，引起它被错误地装载，并在其密码子出现时导入错误的氨基酸。在某些例子中，身份可能由一个碱基对决定（如 $tRNA^{Ala}$）。反密码子只是负责翻译的特异性。如果已填充的氨酰 tRNA 中的氨基酸被化学修饰，修饰残基将被插入到新生肽链中。这样，如同免疫球蛋白，tRNA 具有正常的由通用残基介导的效应功能以及由不同残基介导的特异功能（与氨酰 tRNA 合成酶和 mRNA 的相互作用）。tRNA 中的其他碱基似乎是可变的，其中一些出现在 D 环，但大部分在外臂（见下文）。

二级结构——三叶草模型　tRNA 一级序列中的几个区域的不完善两侧对称性导致四个双链茎的形成。tRNA 的两端配对形成**受体臂**（acceptor stem），包含 3′末端不变的 CCA 基序。在装载过程中，氨基酸与此基序末端的腺苷酸连接，可能与 2′或 3′的羟基作用（参见**氨酰 tRNA 合成酶**）。tRNA 中间的残基折叠形成三个主要的环区：**TΨT 环**，它的得名来自它的固定顺序 GTΨCRANTC（Ψ 是修饰残基假尿苷）；**D 环**，它的得名由于含有修饰残基二氢尿嘧啶；**反密码子环**携带与 mRNA 中的密码子配对的三核苷酸的反义密码子序列。一类 **tRNAs** 中 TΨC 臂与反密码子臂之间相距 4～5残基。在**二类 tRNAs** 可能延伸形成**额外臂**（extra arms），为分子中最大的二级结构，包含 25 个碱基。额外臂的功能仍不清楚。



tRNAs的共同结构:不变与半不变碱基被标出,可变碱基用圆圈表示。下划线表示此碱基往往被修饰,虽然修饰类型不同。Ψ是假尿苷。箭头表示附加碱基可能插入的位置,在D环中包括修饰碱基二氢尿嘧啶。有阴影的碱基在起始tRNAs中很重要,反义密码子茎带有G:C碱基对,受体臂末端不配对。

三级结构　几种转运RNAs的三级结构用X射线晶体衍射确定。分子呈一紧密的L形的构象,TΨT臂与受体臂形成连续的双螺旋,D环与反密码子环形成另一螺旋。这些螺旋相互垂直,TΨT环与D环形成一核心,而受体臂与反密码子环在分子的相对末端形成可移动的臂。

命名法　每一tRNA分子根据其识别的氨基酸用三字母表示来命名,如$tRNA^{ILE}$是携带异亮氨酸的tRNA。同工tRNAs被加上不同的号码,如$tRNA_1^{ILE}$, $tRNA_2^{ILE}$。起始tRNAs的命名不同,在正文中已作了讨论。氨酰tRNA合成酶根据其底物命名,如异亮氨酰tRNA合成酶将异亮氨酸填充到$tRNA^{ILE}$上。它们的命名往往用三字母表示法来简略地表示,即Ile-tRNA合成酶。被填充的tRNA根据同样的原则命名。即如同上面的例子,异亮氨酰$tRNA^{ILE}$, Ile-$tRNA^{ILE}$。

附属功能　除了它们在翻译中所起的重要作用,tRNAs在胞内还承担其他的功能。氨酰tRNAs可作为其他细胞反应的氨基酸给体,在某些*反转录病毒*(参阅)的复制循环中个别tRNAs可作为反转录的引物。在大肠杆菌的His操纵子中tRNA或类似的结构被发现可对翻译进行调节,并在*衰减作用*中(参阅)起重要的作用。

(徐　沁译)

第24章
重组 DNA 与分子克隆

基本概念和定义

- **重组 DNA**(recombinant DNA)是在体外将不同来源的 DNA 分子共价连接而产生的。与构建和应用重组 DNA 有关的技术称为**遗传工程**(genetic engineering)、**基因剪接**或**基因操作**。
- 重组 DNA 技术的基础是一系列使 DNA 能够被精确操作和修饰的关键酶系和技术。使用的主要酶列在表 24.1。最基本的技术包括:①用序列专一性的细菌内切酶(**限制性内切酶**,restriction endonuclease)剪切 DNA,产生确定的 DNA 片段,使用**连接酶**将 DNA 连接;②在**凝胶电泳**上根据分子量大小分离 DNA;③通过**核酸杂交**在复杂混合物中检测特殊的序列;④将 DNA 导入细胞;⑤通过分子克隆或聚合酶链式反应扩增特定的 DNA 分子。
- 在细胞为基础的**分子克隆中**,目标 DNA 片段在体内随增生细胞数量而扩增。缺乏复制起始点(参阅)的 DNA 分子不能复制,所以要由**克隆载体**提供复制起始点,克隆载体是由质粒或病毒衍生的遗传因子,可以携带外源 DNA(**供体、外源、插入或过客 DNA**)。供体和载体 DNA 共价连接,产生**重组载体**(recombinant vector)。它被导入宿主细胞,随细胞生长而复制,产生供体 DNA 的多个克隆拷贝。重组载体是游离的,能够根据它特有的物理化学属性与染色体 DNA 分离,这样可以从大量培养物中重新得到很多被克隆的供体 DNA。体外扩增 DNA 的另一个方法是用聚合酶链式反应,这种方法比分子克隆更快速和容易,且十分灵敏和有力。然而这种方法只能得到相对较短的序列,所使用的酶比宿主细胞中的精确性差,使 PCR 产物产生一些异质性。另外,分子克隆不需要供体 DNA 序列的信息,而 PCR 需要设计引物与特定 DNA 两侧退火。PCR 在本章中不详细讨论(参见 PCR)。
- 插入克隆载体中的 DNA 从初级或二级来源得到。**二级来源**是已经分离的 DNA 克隆,供体 DNA 从克隆中分离以进行进一步操作,这样的过程称为**亚克隆**(subcloning)。在另一方面,**初级克隆**是从原始或**初级来源**中分离供体 DNA,也就是从整个基因组或一大群 cDNA 中得到。尽管偶尔可能从低复杂性来源中直接获得特定的 cDNA 或基因组 DNA 片段,初级克隆经常需要 **DNA 文库**(DNA library)。DNA 文库是克隆在载体上的特定来源的所有 DNA 片段汇集。所需的克隆必须根据供体 DNA 的特殊性质从文库中分离得到,如序列、所编码蛋白的结构或功能特点。这样的过程称为筛选(screening)。

- 一旦特定 DNA 克隆被分离，它可以有很多用途。研究应用属于下面中的一种或多种，在下文中详细讨论：①分析基因和基因组结构及基因表达；②基因物理图谱和定位克隆（见基因结构和作图）；③表达克隆；④基因和产物及调节元件的功能分析；⑤体外突变；⑥基因转移和转基因。

表 24.1 体外 DNA 操作所使用的主要酶，它们的活性和应用

酶	基本性质和应用
Klenow 片段（大肠杆菌聚合酶 I 的一部分，参见复制）	缺乏 5′→3′外切酶活性 用于一般的 DNA 合成，如随机引物标记，末端补平，cDNA 第二链的合成，引物延伸
T4 DNA 聚合酶	有 3′→5′外切酶活性 用于置换标记，产生 DNA 平头末端
T7 DNA 聚合酶	快速和高的进行性 修饰译本，测序酶（sequenase）缺乏 5′→3′外切酶活性，用于 DNA 测序
热温定 DNA 聚合酶（如 *Taq* DNA 聚合酶）	在高温下发挥功能 用于 PCR，并在 DNA 二级结构成为主要问题时使用
反转录酶	RNA 依赖的 DNA 聚合酶活性 用于 cDNA 第一链的合成，RNA 测序
RNA 聚合酶（T7、T3、SP6）	强启动子专一性 DNA 依赖的 RNA 聚合酶 用于体外转录，RNA 标记
限制性内切酶	顺序专一性 DNA 内切酶，有超过 1000 种商品化 用于克隆作图和制备 DNA 片段
S1 核酸酶，绿豆核酸酶	单链专一性 DNA 内切酶 用于 DNA:RNA 杂交体作图，剪切单链 DNA
RNA 酶 A	一般的 RNA 内切酶 用于将 DNA 中污染的 RNA 除去
DNA 酶 I	一般的 DNA 内切酶 在 DNA 中导入缺口，从 RNA 中除去污染的 DNA
T4 DNA 连接酶	产生磷酸二酯键 用于粘端和平端的连接
小牛肠碱性磷酸酶	移去 DNA 和 RNA 的 5′磷酸基团 用于阻止自身连接和 5′末端标记
T4 多核苷酸激酶	在 DNA 和 RNA 加上 5′磷酸基团 用于 5′末端标记
末端脱氧核苷酸转移酶	不依赖模板的 DNA 聚合酶 用于 3′末端标记和加尾

24.1 分子克隆

分子克隆的基本原则　**分子克隆**是体内产生大量特定 DNA 片段的技术。过程中主要有四步（图 24.1）：

(1)构建重组载体，涉及剪切、修饰和供体 DNA 与载体在体外的连接；

(2)将重组载体导入合适的宿主细胞中；

(3)含有载体的细胞选择性增殖（此是克隆阶段）；

(4)抽提和纯化克隆的 DNA。

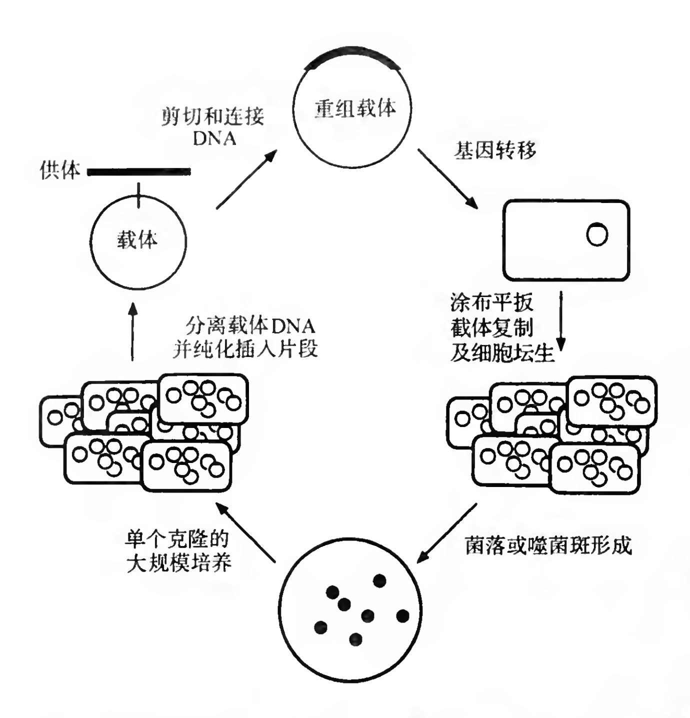


图 24.1 分子克隆的原理。被克隆的供体 DNA 连接到载体中，产生一个重组分子。这个分子被导入宿主细胞，细胞涂布在固体培养基上。重组载体随细胞增生而扩增，克隆可以通过菌落或噬菌斑被鉴定。单个克隆被挑出，接入液体培养，从中可以分离得到大量纯系重组 DNA。

剪切、修饰和连接 DNA 分子　DNA 可以被很多机械、化学或酶学方法随机剪切，处理的程度决定了产生片段的平均大小(表 24.2)，这些方法可以用来产生随机的重叠的基因组 DNA 片段，其中只有用限制性酶才可以产生精确的特定片段(文框 24.1)，细菌内切酶可以识别短的**限制性位点**(restriction site)。大多数 II 型限制性酶识别的限制性位点是二重对称的，在每条 DNA 链的半数识别序列的同样位点剪切磷酸二酯键。少数酶在对称轴上剪切，产生平端的片段，但大多数剪切位点不在对称轴处，所以产生的片段有互补的单链突出端(**粘端或粘性末端**)(图 24.2)。通过氢键，这些片段能与其他同样酶(或有同样专一性的酶)产生的 DNA 片段连接。每个片段末端突出间的碱基配对，随之它们能被 **DNA 连接酶**(DNA ligase)有效地连接起来(图 24.2)。

克隆的最简单策略是用同样的限制性内切酶作用供体和载体 DNA，并用 DNA 连接酶连接。然而这样可使载体在没有插入片段情况下重新环化，产生高背景的非重组载体。有两种方法可防止载体的自我连接：①打开的载体用**碱性磷酸酶**(alkaline phosphatase)去磷酸化——连接需要线性化载体的 5′磷酸基团，如果此磷酸基团被去除，载体不能重新环化除非由供体 DNA 提供 5′磷酸基团；②载体和供体用一对限制性内切酶产生两个不互补的末端——载体不能重新环化除非缺口由同样酶所产生的插入片段来填充。第二种策略略高一筹的是插入片段的方向是可以预定的(**定向克隆**，directional cloning)。

表 24.2　用于克隆的 DNA 来源和对 DNA 分子剪切、修饰和连接产生重组载体的策略

克隆 DNA 的来源
基因组 DNA
cDNA
以前分离的克隆(亚克隆)
PCR 产物
化学合成的寡核苷酸
DNA 剪切机制
通过机械剪切(超声、振荡)、化学处理(如酸/碱水解)或非专一性的内切酶随机剪切→参差不齐的片段
用顺序专一性的限制性内切酶剪切→产生平端或粘端 DNA 片段
末端修饰策略
以 T4 DNA 聚合酶末端补平和修整→平端片段
以 Klenow DNA 聚合酶部分末端补平→可变粘端片段
加上连接区再用限制性酶消化→粘端片段
加上接头→粘端片段
DNA 连接策略
粘性末端的连接
平端连接
同聚物加尾

也见表 21.1 PCR 亚克隆策略。

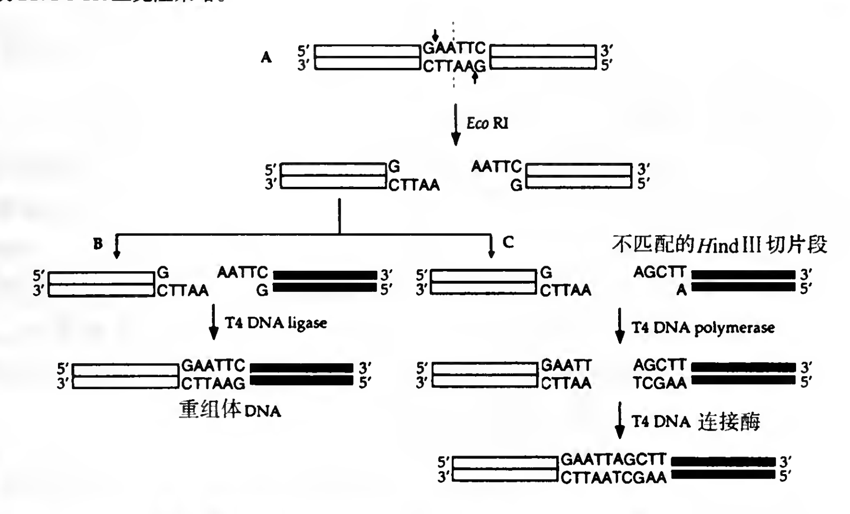


图 24.2　DNA 的剪切、修饰和连接。(a)GAATTC 顺序是限制性内切酶 *Eco*RI 的识别位点。它是二重对称的(即每条链的顺序从 5′→3′方向是相同的),但剪切位置(小箭头)不在对称轴(虚线),所以产生 4 碱基的 5′突出端,也就是粘端。(b)这些能与携带相同末端的外源片段(黑线)相接,使它们能被 DNA 连接酶连接起来产生重组分子。(c)如果有不配对末端的外源片段,一种策略是用 T4 聚合酶使二者末端变成平端再连接它们。其他连接不配对末端分子的策略在文中讨论。

在很多情况下,合适的末端不能有效连接(如供体和载体必须用不匹配的酶切,供体 DNA 是随机剪切,末端是参差不齐,或供体是 cDNA)。在这样的情况下,有其他涉及 DNA 末端修饰的连接策略(表 24.2)。一个常用的方法是通过补平或修整突出末端产生平端。噬菌体 T4 DNA 聚合酶同时有这两个功能,经常用来产生平头末端片段。平头末

端的连接比粘端连接效率低，且没有方向性，但对大多数亚克隆来说这样的效率已是足够了。另一种策略是在供体或载体 DNA 末端上加连接体或接头。**连接体**(linkers)是双链寡核苷酸，在这种情况下，含有限制性酶切位点。它可以连接到平端供体 DNA 上，再用限制性内切酶作用产生适合连接的粘端片段。接头(adaptors)是已经限制性酶消化后的连接体。将一个接头连接到平端供体 DNA 产生一个不需要限制性酶处理的适合连接的片段；这在当供体 DNA 内部含有所使用的限制性酶切位点时经常使用。将供体和载体 DNA 连接的另一种可能的方法是通过**同聚物加尾**(homopolymer tailing)。平端或参差不齐的供体 DNA 通过末端脱氧核苷酸转移酶(参阅)在每条链的 3′末端加上核苷酸序列(加尾)。如果在反应中只加入一种核苷酸，则形成同聚物尾巴。在线性化载体每条链的 3′末端可以加上互补的同聚物尾巴。当两者混合，尾巴间的碱基配对形成松弛的环。这样的形式可以被转化到细菌中，在体内被修复。

克隆载体　分子克隆涉及在宿主细胞中通过复制来扩增供体 DNA。然而，因为供体 DNA 一般缺乏复制起始点，它必须与一个合适的复制子连接，以有利于克隆。这样的复制子称为**克隆载体**(cloning vector)，它是质粒、病毒或染色体的衍生物。理想的克隆载体应有如下属性：

(1)它们是游离的，也就是说它们并不整合到宿主基因组中，所以它们能够较容易地从大量染色体 DNA 中被分离出来；

(2)它们能自主复制，达到很高的拷贝数，这样有利于所携带的任何外源 DNA 的扩增；

(3)它们使携带载体的细胞从那些缺乏载体细胞中选择出来；

(4) 它们使携带载体的细胞能从那些带有非重组载体的细胞中被选择出来；

(5)它们有通用性，有便于插入的单一克隆位点和有利于克隆后下游操作的辅助功能。

自然的质粒和噬菌体不是好的载体，因为它们缺乏克隆的通用性。发展载体的重要起始步骤是构建含单一限制性酶切位点(**克隆位点**)的质粒，这样使供体 DNA 能插入特定的位置，而不丢失载体顺序。理想的情况是这些位点处于一个可选择或可见的标记基因中，通过插入失活有利于重组体的筛选。目前大多数载体有**多克隆位点**(multiple cloning sites)或**多位点接头**(polylinker)，它们是一簇单一的限制性酶切位点，为用不同内切酶制备的供体 DNA 导入提供了方便的位置。

不同的载体系统用来克隆不同目的不同大小 DNA 片段(表 24.3)。一般目的质粒载体用来亚克隆和下游操作，因为它们分子量小，易于操作，且非常通用。质粒载体携带标记用来选择转化细胞，能在选择性培养基中形成菌落。噬菌体 λ 载体用来构建 DNA 文库，因为它们比质粒容量更大，在长期保存中更稳定，且噬菌斑比菌落更容易筛选。携带噬菌体的细胞筛选有所不同，噬菌体本身能够溶解细菌细胞，在细菌菌苔上形成噬菌斑，噬菌体是根据它的这一特性筛选的。很多载体同时拥有噬菌体和质粒的特性。**粘粒**(cosmid)是携带 λ*cos* 位点的质粒，能够被包装在噬菌体头部。基本的粘粒非常小，所以这些载体能够装载很大的供体 DNA 片段，用于基因组文库构建和重叠群作图(contig mapping)(参阅)。**噬粒**(phagemids)是携带 M13(或类似)噬菌体的复制起始点的质粒，这样在以反式提供合适的噬菌体功能时，使质粒能够作为单链 DNA 噬菌体复制。这样产生单链 DNA 的载体用于诸如测序、体外诱变和探针合成。最后**噬粒**是复合的 λ 质粒载体，λ 插入载体基本上含有整个质粒。这样的载体，如 λZAP，非常通用，允许在噬菌体载体中构

建 cDNA 文库，作为质粒切离，不需要亚克隆，易于下游操作。

表 24.3 不同克隆载体系统的基本性质和应用

一般目的克隆载体

质粒载体

基础：在多拷贝松弛型质粒中自然发生。大多数现代的质粒载体是基于 ColE1 复制子(见第 20 章质粒)

导入宿主：通过化学辅助转化、电转化或脂质体转化裸露的质粒 DNA

载体筛选：载体携带显性选择标记基因，也就是说抗生素抗性标记，使含载体的细胞能在选择性培养基上形成菌落

重组筛选：经常用选择性标记或可见标记的插入失活

供体 DNA 的大小：0～20kbp，尽管 5～10kbp 的插入片段不稳定

主要的应用：亚克隆和下游操作、cDNA 克隆和表达

评论：大多数质粒载体是高拷贝载体，有利于供体 DNA 的扩增。偶尔克隆的 DNA 可能影响宿主细胞，如消耗宿主编码的转录因子等，一些**低拷贝载体**(low copy number vector)设计用来克隆引起剂量敏感的序列(**有害序列**)。一些低拷贝载体提供复制的诱导控制，当细胞数量增加足够时，质粒合成可以爆发(**失控载体**)。质粒载体进行了很大改进以有利于下游应用。**转录载体**由噬菌体含侧翼多克隆位点的启动子组成，通过体外转录产生有义或反义 RNA。大肠杆菌启动子能被用于促进蛋白的过量表达，在上游基因片段的读框中插入，能产生融合蛋白和触发蛋白的分泌(参见*表达克隆*)。加入报告基因(参阅)可用于调节元件的分析。来自其他物种的复制起始点可在大肠杆菌和其他宿主中复制和维持(参见*穿梭载体*)。另见*噬粒*及下文

噬粒载体

基础：质粒复制子加上噬菌体 M13(或类似的)复制起始点

导入宿主：与质粒一样

载体筛选：与质粒一样

重组体筛选：与质粒一样

供体 DNA 的大小：与质粒一样

主要应用：产生单链 DNA

评论：起先，如 M13 和 f1 单链 DNA 噬菌体是用来产生单链重组 DNA，进行测序，探针合成等。这些噬菌体是很简单的载体，缺乏不必要基因，即使中等大小的插入片段也不很稳定。噬菌体的复制起始点使质粒由一个助噬菌体以反式方式提供噬菌体功能时作为单链 DNA 噬菌体基因组复制。噬粒将质粒的通用性与噬菌体产生单链 DNA 的能力结合起来

λ 噬菌体载体

基础：λ 噬菌体

导入宿主：裸露的噬菌体 DNA 的转化或体外包装重组噬菌体进入 λ 衣壳，再将供体 DNA 通过自然感染途径转导

载体筛选：在细菌菌苔中形成噬菌斑

重组筛选：插入载体——通过可见标记的插入失活(*cI* 基因的阻断阻止溶原性，产生的噬菌斑更清晰；现代的载体携带△*lacZ* 基因，可进行蓝-白筛选)。置换载体——Spi^- 表型(对 P2 感染敏感性消失)是中心“填充片段”*gam* 和 *red* 基因座丢失引起的。插入大片段的正选择(如果基因组野生型的序列小于 75%，λ 噬菌体不能形成噬菌斑)

供体 DNA 的大小：插入载体——0～10kbp；置换载体：9～23kbp。插入大小范围被有效形成噬菌斑的野生型基因组 75%～105%大小所限制

主要应用：插入载体——cDNA 克隆和表达文库；噬菌体展示。置换载体——基因组 DNA 克隆

评论：有两类 λ 载体。**插入载体**有单一限制性酶切位点，可将小片段 DNA 加进 λ 基因组。噬粒(参阅)是含有质粒的插入载体。**置换载体**有配对的位点限定的特点填充片段，填充片段含有溶原和重组所需基因，不需要裂解周期的基因。填充片段可以被去除，供体 DNA 在臂间插入

粘粒载体

基础：含有 λ 噬菌体 *cos* 位点的质粒

导入宿主：在体外将重组粘粒包装进 λ 衣壳，再通过正常的感染途径转染

载体筛选：类似于质粒

重组筛选：可见标记插入失活和小片段插入的正选择

供体 DNA 大小：30～45kbp。插入大小范围被有效形成噬菌斑的野生型基因组 75%～105%所限制。

主要应用：基因组文库构建

评论：粘粒使用 λ 包装位点进行有效的宿主细胞释放和大片段筛选。粘粒广泛用于基因组作图和重叠群排列

大容量克隆载体
细菌人工染色体[(bacterial artificial chromosomes, BACs,养粒(fosmids)]
基础:大肠杆菌 F 质粒
导入宿主:电转化
载体筛选:显性选择标记
重组筛选:插入片段大小
供体 DNA 大小:>300kbp
主要应用:大基因组分析
评论:低频率的重排和嵌合分子。载体在每个细胞中维持一个或两个拷贝,产生少量供体 DNA

P1 载体和 P1 人工染色体(P1 artificial chromosomes, PACs)
基础:噬菌体 P1
导入宿主:体外包装和转导
载体筛选:显性选择标记
重组筛选:有多种——在一个系统中,应用于对致死标记的阻断进行正选择
供体 DNA 大小:大约 100kbp
主要应用:大基因组分析
评论:低频率的重排和嵌合分子。载体维持低拷贝,但可以通过诱导噬菌体 P1 产生裂解周期而扩增

酵母人工染色体(yeast artificial chromsomes, YACs)
基础:酿酒酵母着丝粒、端粒和自主复制序列(染色体复制起始点)
导入宿主:转染酵母原生质体
载体筛选:显性筛选标记(营养缺陷型的恢复)
重组筛选:插入片段大小
供体 DNA 大小:>2000kbp
主要应用:大基因组分析,YAC 转基因小鼠(参阅)
评论:YAC 是高容量克隆载体,但有一些缺点,包括高频率的自发缺失和克隆的嵌合化。重组载体的大小因分辨率的原因,需要电泳分析,有时难以将 YACs 从内源性酵母染色体中分离。YAC 维持在低拷贝(经常一个细胞中只有一个)

最后,大的真核生物基因组分析需要发展高容量的载体——**人工染色体**——用于基因组作图和大基因和基因复合物的结构功能的分析。酵母人工染色体是这些载体中的最完善的形式,拥有最大的容量,但也有高频率的**克隆嵌合化**(不相关供体 DNA 片段的连接和维持)。最近,很多基于细菌质粒的人工染色体载体得到关注。它们比 YACs 容量小一些,但嵌合插入少很多。

DNA 导入克隆宿主　一旦在体外构建好重组载体,它必须被导入宿主细胞进行克隆。大肠杆菌是一般目的克隆的主要宿主,但这种细菌自然情况下不具有从周围培养基中吸收 DNA 的能力。可以通过化学处理产生有这种能力的人工状态,如与二价阳离子一起孵育。一个短暂的热休克刺激吸收 DNA,在适宜的条件下每 μg 载体 DNA 产生 $10^{7\sim9}$ 细菌菌落或 $10^{4\sim5}$ λ 噬菌斑。细菌吸收质粒 DNA 的过程称为**转化**(transformation),吸收裸露的噬菌体 DNA 为转染(transfection),尽管两种情况的机制是一样的(转化和转染在应用到真核细胞时是不同的途径,参见表 24.11)。**电穿孔(电转化)**是另一种技术,DNA 通过短暂的高电压产生的孔进入细胞。这也是细胞导入 DNA 的有效方法,每 μg 载体 DNA 可以产生 10^9 的菌落。这对如 BACs 这样的低拷贝质粒尤其重要。

尽管这些技术对亚克隆已经足够了,但 λ 噬菌体文库构建需要更高效的 DNA 转移。噬菌体和噬粒载体能够通过**转导**(transduction)被高效转移,这涉及先将载体包装在 λ 噬

菌体的头部(**体外包装**)。这通过重组载体 DNA 与噬菌体头前体、尾巴和包装蛋白混合,然后以包装好的克隆感染细菌进行培养。DNA 通过正常的噬菌体感染途径进入细胞,每 μg 载体 DNA 可以产生 10^6 个菌落(噬菌粒)或噬菌斑(λ 噬菌体)。

DNA 导入酵母细胞在文框 24.4 中讨论,DNA 导入动物和植物细胞在表 24.9[1]中讨论。

载体和重组筛选　　无论是 DNA 操作还是基因转移过程都不是 100%有效的。这样在任何克隆实验的开始,都会有大量缺乏载体和含载体的细胞,中度比例的细胞含非重组的载体。空细胞和非重组细胞可在以牺牲重组群体为代价的情况下增生,所以希望鉴定和去除这些细胞。

载体筛选是对携带载体细胞的筛选——经常是正的和直接筛选,也就是说载体具有或赋予细胞一种可筛选的属性。转化质粒载体的细菌细胞通过质粒携带的显性抗生素抗性标记进行阳性筛选,有效地维持含质粒细胞的群体。在真核系统中用另一种标记,如营养缺陷型(参阅)的恢复经常用于酵母中筛选 YACs,尽管这需要特殊的营养缺陷型酵母菌株。对噬菌体载体而言,噬菌体本身可以通过在细菌菌苔中裂解细胞形成噬菌斑的能力来筛选。

重组筛选是在携带非重组载体细胞中选择含重组载体细胞。根据载体类型有很多不同的筛选系统可以应用。重组质粒经常通过插入灭活第二个标记,或第二个抗生素抗性标记(在筛选步骤中需要影印培养的过程)或一个可见的标记。目前普遍的策略是**蓝-白斑筛选**。质粒携带一个无功能的,截断的 *lacZ* 等位基因,编码 β-半乳糖苷酶的 N 端的小片段,称为 **α-肽**。这可以通过编码多肽其余部分的等位基因得到互补,这在如 JM101 等特别修饰的宿主菌中存在(**α-互补**)。有功能的 β-半乳糖苷酶将无色的底物 X-gal(见表 24.7)转变成蓝色沉淀。△*lacZ* 基因含内在的多接头区,供体 DNA 插入造成失活。在正确的检测培养基中,从白色菌落中得到重组细胞和蓝色菌落中为非重组细胞。一些直接负选择质粒是根据第二个标记基因的条件致死性设计的,使含非重组质粒的细胞在根据它们对标记失去敏感的特性,在限制性条件下被反筛选。然而,很多这样的载体需要少有的特殊宿主菌株和筛选体系。

重组 λ 插入载体经常是可见选择(或阻断 *cI* 基因——阻止溶原性,产生更清晰的噬菌斑——或阻断上面所讨论的△*lacZ* 基因)。重组 λ 置换载体根据它们能感染那些对 P2 噬菌体溶原化的细菌和供体 DNA 的大小具有两种正选择。野生型的噬菌体有 Spi^+ 表型,因为它们对 P2 感染敏感,也就是说它们不会感染已经被噬菌体 P2 感染的细胞。这种敏感性是由于填充片段的 *gam* 和 *red* 基因座提供的,填充片段可以被供体 DNA 切除和置换。因此只有重组载体才能在 P2 溶原中形成噬菌斑。另外,λ 只有在重组基因组是野生型基因组大小的 75%～105%时,才能形成感染颗粒。105%的上限表明了插入和置换载体的最大插入片段大小。非重组插入载体的基因组大约是野生型的 100%,而缺乏填充片段的非重组置换载体由于低于 75%,不能形成噬菌斑。同样的原则适用于粘粒载体,因为它们在噬菌体头部中进行包装:载体筛选像常规质粒载体一样依赖于抗生素抗性,但重组筛选依赖于类似标准 λ 载体的插入片段大小。

---

1) 原文误写为表 24.10。——译者注

并不总是需要对重组载体进行筛选的。在一个简单的亚克隆实验中，连接反应可以被控制，只产生很少量的非重组体背景，这样随机挑选菌落得到所想要的克隆的可能性很大。同样，当菌落或和噬菌斑是用来杂交分析时，非重组体不能与探针杂交，在进一步分析中可以被去除(参见菌落筛选，噬菌斑挑出)。

克隆的 DNA 的恢复 在重组 DNA 导入宿主细胞后，细胞培养较短时间就能恢复和涂皿(在固体培养基上涂布)，在合适的选择条件下形成菌落或噬菌斑。经常，每个菌落或噬菌斑代表了相同细胞或噬菌体的克隆，能够被**挑出**(从平板上移出，转移到)在液体培养基中进行第二轮克隆(这次从其他克隆中分离)。平板接种是对异源重组载体群体分级分离成同源克隆，在平均水平上，每一个宿主细胞有一种重组分子。平板接种还是 DNA 文库筛选的关键步骤，在这过程中，每一个菌落或噬菌斑代表了基因组的不同部分或不同的 cDNA。

分子克隆的最后步骤是所克隆 DNA 的恢复。传统的方法涉及细胞裂解，及一系列选择性沉淀、沉降分离步骤，是费时费力费钱的。最近的创新包括了通过玻璃珠或树脂对 DNA 的吸附。很多商品化的方便试剂盒能在几小时内从细胞中得到大量纯的质粒或噬菌体 DNA。

24.2 基因分离的策略

从简单和复杂来源中分离 DNA 片段 上面讨论一些使任何 DNA 顺序插入载体和克隆化的技术，以有利于进一步的分析和操作。在来源 DNA 不复杂的情况或特定顺序高度富集的情况，直接分离所要的供体 DNA 片段，将它插入载体进行克隆是可能的。这在来自小基因组中基因组 DNA 片段（例如质粒、一些病毒和动物线粒体）和已经获得的克隆、PCR 产物和特定组织中代表性的高度富集的 cDNA(如珠蛋白、卵清蛋白)中有应用。

在更多的情况中，特定目标序列的来源是很复杂的(例如人类基因在基因组 DNA 中平均是百万分之一)。这就需要构建一个 DNA 文库，汇集特定来源的所有克隆在载体中 DNA 片段。文库主要有两类：从所有基因组 DNA 中制备的**基因组文库**(genomic libraries)和由 mRNA 分子反转录制备的 **cDNA 文库**(cDNA libraries)。在多余序列的背景中(通常是大量的)需要通过称为筛选(screening)的过程鉴别出感兴趣的序列。根据所提供的信息选择筛选方法(表 24.4)。

基因组文库 基因组文库的大小(代表整个基因组所需要的单个克隆的数量)不仅依赖于供体 DNA 片段的平均大小和基因组的大小，而且依赖于基因组中特定区域出现的概率。基于每个顺序出现一次的原则，文库构建并不简单。因为不同的克隆效率和载体构建和基因导入宿主细胞时取样误差，而存在有些序列出现不止一次而有些根本不出现的情况。也有一些重叠的片段的存在，这样有利于排列克隆重叠群而产生完整的物理图谱和基因序列(见基因结构和作图)。平均大小适当的重叠片段可以通过基因组 DNA 的随机剪切或用四碱基限制性内切酶的少量消化(这些酶切位点在 DNA 中有很多，分布非常随机，通过在这些位点罕有的剪切产生随机和重叠的片段)。后一种策略更方便，因为片段末端是粘性的，供体 DNA 可以直接与载体连接。

表 24.4　根据所提供的来源和信息，从 cDNA 或基因组文库中分离基因的策略

所提供的信息	筛选策略
功能克隆——不需要表达	
转录本在特定来源组织中高度富集	丰富克隆——从 cDNA 文库中随机选择克隆，测序确定产物
部分核苷酸顺序已知	用寡核苷酸探针筛选[a]
可用的部分克隆(如用 cDNA 筛选基因组文库)，或根据同源性筛选相关克隆序列	在低严格条件[a]下用克隆的片段或同源基因筛选[a]
已知部分多肽序列	用简并性寡核苷酸[a][**猜测体**[a](guessmers)][1)]筛选文库
两种组织中差异表达	加减筛选 通过消减杂交[a]可得到富含差异表达克隆的丰富文库(参见*差异克隆*)
涉及 cDNA 表达的功能克隆	
提供突变	通过对突变表型的互补筛选(**表型恢复**)
提供抗体	用免疫检测筛选文库
特殊特性的产物	通过特殊技术筛选文库，如对 DNA 结合蛋白的 southwestern 杂交，相互作用蛋白的*酵母双杂交体系*(参阅)，通过酶转化底物
定位克隆	
提供结构突变	如果突变由缺失引起，从基因组消减文库克隆(参见*差异克隆*)
由转座子插入引起的突变	如果突变由转座因子的插入而引起，用转座子序列作为探针筛选突变体文库，通过质粒拯救分离克隆(参见*转座子标记*、*质粒拯救*)
基因在染色体上的位置	通过与标记或染色体断点相连接的染色体步移(chromosome walking，参阅)定位克隆，标记可能是增强子包载载体(参见)的转座因子
非筛选方法	
基因组作图和测序	对跨越整个基因组的克隆进行系统化分析(见基因结构和作图)
表达序列标记	对随机 cDNA 克隆的大量克隆和分析(参见*表达序列标记*)

a　以 PCR 为基础的方法也可应用(参见第 21 章 PCR)。

下面的公式用来预测克隆的总数目 N，能以概率 p 获得达到文库包含了所有序列的可能性，n 是覆盖基因组一次理论上要得到的克隆数目，也就是**基因组当量**(genome equivalent)：

$$N=\frac{\ln(1-p)}{\ln(1-1/n)}$$

预期序列有 95% 的可能性包括在文库中需要预测 3～4n 的克隆，而达到 99% 的可能性(经常是最佳标准)，文库必须包含 4～5n 克隆。构建在 λ 载体上的一个大肠杆菌基因组文库(平均插入大小为 20kbp)需要 800～1000 个克隆，而相对的人的文库需要大于 0.5 百万克隆才能达到同样的概率含有特定的基因。

文库大小可以通过用构建在大容量载体(如粘粒或人工染色体)上的大插入片段作为起始筛选来降低。如果所待测基因染色体位置已知，有可能通过分离单个染色体并构建**染色体专一性基因组文库**(chromosome-specific genomic libraries)，从中进行克隆。染色体分离可以通过**荧光激活的染色体分选**(fluorescence-activated chromosome sorting, FACS)达到，分离染色体是根据染色体与特定染料的不同结合能力，或用*单染色体体细胞杂交*(参阅)。同样有可能产生染色体特定区域的文库，或是用含缺失的染色体作为文库构建的材料，或是通过**染色体显微解剖**(chromosome microdissection)。

cDNA 文库　cDNA 是**互补 DNA**，也就是说与 mRNA 互补的 DNA。cDNA 文库是通过

1)　用于基因克隆的低简并性寡核苷酸探针，其序列按已知的氨基酸序列推导，但仅采纳据猜测最可能与目的基因配对的密码子。——译者注

反转录的 mRNA 群体，并制备**双链 cDNA**(double stranded cDNA)来构建的。这样 cDNA 文库在顺序表现度、基因序列结构和所使用的筛选方法上与基因组文库不同：

(1) cDNA 文库代表了 mRNA 的来源，其中一些特定的转录本丰富而另一些很少。这样不像基因组文库，理论上均等地代表了所有基因序列，cDNA 文库对一些顺序相对丰富而另一些缺乏。cDNA 文库可用来分离丰富的 cDNAs，而另一方面分离稀少转录本的 cDNAs 是困难的。从不同细胞类型(或不同发育阶段，或不同处理的细胞)制备的 cDNA 文库会包含一些共同序列和一些独特序列。这可被用于分离差异表达的基因(参见*差异克隆*)。

(2) cDNA 文库代表的是表达的 DNA，它们缺乏内含子、调节元件和基因间 DNA。在研究基因结构和调节中 cDNA 文库用处不大，但它一般比基因组克隆要小，且真核生物 cDNA 可以在细菌中表达(它们不能剪接内含子)。不同的剪接产生不同的但部分重叠的 cDNA 克隆。

(3) 基因组文库通过杂交筛选。而因为 cDNA 可以在细菌中表达，*表达文库*(参阅)可用于多种筛选策略，如免疫筛选和互补筛选(表 24.4)。

制备一个 cDNA 文库(图 24.3)涉及分离代表细胞中主要 mRNA 的 poly(A)$^+$ RNA (rRNA 和 tRNA 因其丰富和大小的一致性，可以通过分级分离直接分离)。**Poly(A)$^+$ RNA** 的选择可以通过多抽提核糖体得到或者更经常的是将细胞总 RNA 经过一个基质上含有多尿嘧啶或多胸腺嘧啶的柱子(亲和选择)来分离。纯化的 poly(A)$^+$RNA 然后被反转录。一个共同的策略是用 oligo(dT)引物引发 **cDNA 第一链的合成**用，尽管这样最终文库倾向于产生 3′末端，因为长的 mRNA 没有被完全反转录。另一种策略是使用随机六核苷酸引物，在所有 RNA 的随机位置退火，产生重叠的 cDNA 片段，最终可以组建成全长 cDNA。**第二链 cDNA 的合成**是可被自身引发的，因为第一链倾向于折回形成发夹。这可用于 cDNA 的定向克隆(图 24.3)，但会丢失一些 5′顺序。一个更有效的方法是在 cDNA 第 一链加尾(如多聚胞嘧啶)，然后用oligo(G)为引物起始第二链的合成。这样产生的双

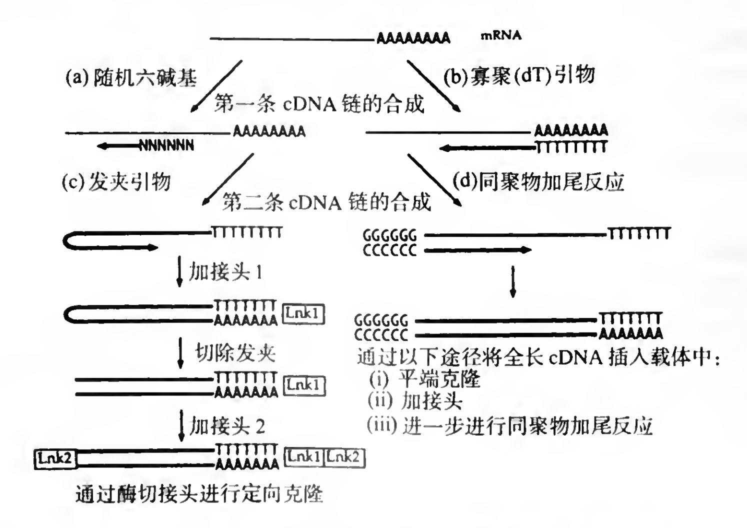


图 24.3 cDNA 克隆。cDNA 第一链的合成是用(a)随机六碱基或(b)寡聚(T)引物通过反转录 poly(A)$^+$RNA。第二链的合成可以用(c)自身引物或(d)第一链的加尾的互补引物。自身引物由于核酸酶的处理会失去一些 5′顺序，但使定向克隆成为可能。DNA 以粗线表示，RNA 以细线表示。

链 cDNA 可以通过加接头连接物或同聚尾巴插入载体。

筛选策略　经典的筛选策略需要基因的生化知识或一些序列信息，或者利用产物的性质能够使基因在表达文库中分离。这些应用列在表 24.4 **功能克隆**(function cloning)中，不需要任何关于基因的位置情况。相对的方法是**定位克隆**(positional cloning)，是无需任何生化或功能信息，但要了解染色体图谱中大致位置的基因分离。

尽管功能克隆和定位克隆的策略都是有用的，但基因分离还是艰苦的工作，它们正被最近发展的工厂式**非筛选策略**(nonscreening strategies)所取代，这种策略涉及系统化克隆和分析基因组 DNA 和 cDNA。这已被基因组计划中出现的知识和人类表达序列标记(参阅)的分析所证实(见基因结构和作图)。

功能克隆　功能克隆策略可以分为三个方面：①利用序列信息——这不需要克隆基因的表达；②利用基因表达的信息——这不需要克隆基因的表达；和③利用蛋白功能——这需要克隆基因的表达，只适用于 cDNA 文库。

传统的序列为基础的方法包括用部分序列信息通过杂交分离克隆(见文框 24.3)。一个简单的例子是用 cDNA 克隆分离相关的基因组克隆或反之亦然。另一种情况，克隆可以通过已经克隆的与其有同源性但不完全一样的序列(如同一个多基因家族中的另一个成员，或不同物种中的类似基因，或高度保守的组件，如同源异形盒)进行分离。如果多肽序列已知，寡聚核苷酸探针可以根据对应的核酸序列设计，这种情况存在一些简并性(参阅)因此必须考虑到在探针设计中编码同样肽段的所有可能的密码子。

利用基因表达信息，而不需要克隆确实被表达的方法，是用特定 cDNA 克隆表现度的差异，反映了来源材料中对应 RNA 的丰富度。这样的方法包括特定文库中极其丰富基因的克隆(**富集克隆**)和利用文库和 RNA 来源差异来分离差异表达的基因(参见差异克隆)。

建立在 cDNA 表达上的筛选策略利用编码多肽的结构和功能。这些方法中最直接的是对表达文库的免疫筛选，用标记的抗体鉴定表达特定多肽的克隆。类似的不同的其他分子也可以作为探针，包括检测核酸结合蛋白的 DNA 和 RNA 及检测蛋白-蛋白的相互作用的蛋白。蛋白的活性也可以被筛选，如用进行特定反应的能力或测定突变表型的恢复来测定编码酶的克隆。

定位克隆　如果基因的大致位置知道，它可以在需要序列或生化功能知识的基础上被克隆。定位克隆从鉴定附近已经克隆的标记开始。目标基因近的标记(如人类遗传疾病的已知相关基因)可以通过连锁分析或低分辨率物理作图来确定(参见基因组和作图)。

在人类和其他动物，定位克隆的第一步是**染色体步移**(chromosome walk)。这包括用邻近的标记作为探针在基因组文库中筛选重叠克隆。阳性克隆可作为下一步的探针，直到作图基因区域被覆盖，一系列邻近克隆跨越了感兴趣的区域。步移的方向和进度可通过中期染色体的原位杂交确定。这种技术在原理上很简单，但大范围的染色体步移有技术上困难，因为有重复 DNA 或不能克隆顺序(在文库中不存在的顺序)存在，这样的障碍通过**染色体跳查**(chromosome jumping)能有效地克服：基因组 DNA 的大片段克隆在粘粒载体中，这样原来在基因组中距离很多 kbp 远的 DNA 顺序当载体闭合时被带到一起。邻近位置的 DNA 可以被亚克隆产生**跳跃文库**(jumping library)，跳跃文库快速沿染色体进行。重复 DNA 的效应通过与 **Cot-1 DNA** 的竞争性杂交得到相当程度控制，Cot-1 DNA

对应于卫星 DNA 和超高丰度因子，在复性实验（参见 *Cot* 分析）中迅速退火。

一旦作图的基因座被克隆，有很多技术可用于在区域中鉴定**候选基因**（candidate genes）（见表 12.10）。然而随作图和克隆基因数目的增加，通过扫描已被步移确定覆盖区域的基因组数据库来选择一组候选基因的可能性大为增加。一旦一组这样的候选基因被鉴定，就需要通过受影响的个体有突变，而不受影响个体没有突变，来确定它们中哪一个是目标基因。一些**突变筛选**的方法在表 24.5 中列出。

在植物中，大量重复 DNA 使染色体步行不能进行，但在很多物种中以 RAPDs（参阅）为标记作出的高密度物理图谱的有效性意味着这些标记对于那些基因常常靠得很近，足以将它们包含在同一基因组克隆中。因此，这些过程使得一个适当的标记可直接用作扫描基因组文库的探针，此过程被命名为**染色体登陆**（chromosome landing），与转座子标记技术（参阅）类似。

表 24.5 鉴定克隆 DNA 突变的突变筛选

技术	检测的基础
在整个基因组中检测不同类型的突变技术	
细胞遗传学作图	通过细胞遗传学鉴定染色体畸变。可以显示表型和特定重排间的联系
比较基因组杂交	竞争性原位杂交的形式，鉴定不同扩增的基因。在检测引起癌症的扩增突变中非常有用（见第 18 章癌基因与癌）
确定特定 DNA 片段中点突变和插入及缺失突变定位的高分辨率技术	
RFLP 分析	分析克隆 DNA 限制性酶切片段长度多态性（参阅），多态性由限制性位点的产生/缺失（点突变）或重复扩充/收缩（插入和缺失突变）引起
变性梯度凝胶电泳（denaturing gradient gel electrophoresis，DGGE）	双链 DNA 在增加变性剂的梯度凝胶中迁移，直到标准分子分离。不同等位基因的同源双链和异源双链在细微差别条件下变性，产生不同的条带
持续变性毛细管电泳（constant denaturant capillary electrophoresis，CDCE）	不同等位基因的同源双链和异源双链在一定变性条件下，部分变性时在胶中有不同的移动能力
单链构象多态性分析（single-strand conformational polymorphism，SSCP）	DNA 单链片段由 PCR 扩增并在非变性胶中分离。点突变影响单链所采用的四级结构和它们的迁移。这种技术不能精确地确定突变位置，但因为它只适用于短片段（<300nt），所以在基因中能作突变扫描
限制性内切酶指纹（restriction endonuclease fingerprinting）	更精炼的 SSCP 分析，从不同等位基因扩增的 DNA 用限制性酶消化产生含突变的不同大小片段，至少其中一个显示构象多态性。突变位置可以通过限制性图谱（参阅）确定
蛋白截断测试	与cDNA 克隆的体外转录和翻译相连，用电泳分析蛋白。检测由缺失/重排和点突变引起的蛋白截断
DNA 片段中检测点突变的精确位置的技术	
错配的化学剪切	异源双链 DNA 在错配的胞嘧啶（通过哌啶和羟胺的修饰）和胸腺嘧啶（由四氧化锇）被特异剪切，片段在测序胶上通过电泳分离确定错配位置
酶错配剪切	异源双链 DNA 被 T4 内切酶 VII 剪切，接着如上进行电泳分离和分析
错配修复酶剪切	通过错配修复酶对异源双链 DNA 的剪切。例如大肠杆菌 MutY 检测 A:G 和 A:C 错配。其他酶可用于检测不同的错配碱基。接着如上进行电泳分离和分析
测序比较	测序，直接鉴定插入和缺失突变和错配位置

注：突变筛选用于鉴定未知突变，突变检测（参阅）用于鉴定已知突变。很多高分辨率的筛选技术是利用异源双链 DNA，不是作为错配确定突变就是通过它们与同源双链在电泳胶上不同的迁移作比较。异源双链 DNA 来源反映了共同的从杂合子中扩增 DNA 库的策略，这种策略是用 PCR 在热循环中随机与野生型和突变链相联系。

差异克隆　**差异克隆**是利用来源不同 DNA 之间的表现度差异来分离差异表达基因的功能克隆方法。在任何组织中主要表达的基因是管家基因，在大多数 cDNA 文库中都出现。为了专一性触发独特的基因，称为**差异筛选**(differential screening)和**加减筛选**(plus and minus screening)技术，包括用两个不同组织来源的标记 mRNA 或 cDNA 筛选两个重复文库。从一个文库来源的标记 mRNA 与大多数克隆杂交，而从第二个来源的标记 mRNA 只和部分克隆杂交，这些代表了共同的管家基因。被一个探针鉴定而不与另一个杂交的克隆是差异表达的，可以分离以进行进一步分析。

一个更有效的鉴定差异表达基因的途径是构建独特克隆丰富的文库。这可以通过**扣除杂交**(subtractive hybridization)得到，用一个来源的 cDNA 去除第二个来源中共同的 RNA，留下独特的 RNA 用于文库的构建。对生物素标记的 cDNA 与亲和素亲和捕捉(文框 24.4)对发展这项技术是非常有价值的。一种来源的 cDNA 是用掺入生物素标记核苷酸制备的。cDNA 通过变性、与第二个来源的 mRNA 混合杂交。如果应用过量的标记 cDNA，大多数的共同 RNA 会形成 cDNA:RNA 异源双链，与 cDNA 同源双链和单链一起，通过固定化的亲和素玻璃珠的亲和吸附分离。经过几轮这样的消减(加入新鲜的 cDNA)，保留的 mRNA 是独特转录本的高度富集，可用于构建**消减 cDNA 文库**。

同样的策略可用于在基因组文库中定位克隆那些根据染色体缺失分析确定的基因。如果基因组 DNA 能从正常和突变(缺失)的来源中制备，消减杂交能够富集跨越缺失区域的基因组克隆。这种技术成功地克隆了杜兴氏肌营养不良(Duchenne muscular dystrophy，DMD)基因。

不同的克隆可以通过以 PCR 为基础的技术获得(参见表现度差异分析、差异显示，抑制 PCR)。

24.3　克隆 DNA 的性质

限制性图谱　一旦分离得到一个新的克隆，分析的第一步就是做**限制性图谱**，限制性图谱是不同限制性内切酶相对位置的物理图谱。载体以一系列限制性内切酶消化，限制性片段用凝胶电泳(参见文框 24.2)分离。每个限制性位点的位置通过单酶切、双酶切和部分酶切产生的片段大小来确定(一个例子显示在图 24.4 中)。与通过电泳的简单分析一样，末端标记的供体 DNA 确定图谱的边界，单个的限制性片段可以被标记，通过 Southern 杂交(文框 24.3)鉴定重叠的片段。限制性图谱显示了供体 DNA 的大小，并可鉴定用于亚克隆的位点。产生的限制性片段的带型可用来鉴定用于重叠群作图的重叠克隆(参见限制性片段指纹)。

像 *Eco*RI 和 *Hin*dIII 这样的标准六碱基剪切酶在质粒和噬菌体克隆作图中非常有用，因为这样的限制性位点平均 4kbp 出现一次。对于像粘粒和人工染色体这样的大载体要用稀少的切点(文框 24.1)，此被称为**大尺度限制酶切作图**(long range restriction mapping)需要特殊的电泳方法来分离所产生的大片段(文框 24.2)。

克隆 DNA 的最终分析是确定它的核苷酸顺序。这对于大的克隆来说是不可能的，它必须被亚克隆来提供直接测序所需的小片段。DNA 测序技术和它在基因组分析中的作用已经被重视起来(见文框 12.2)。

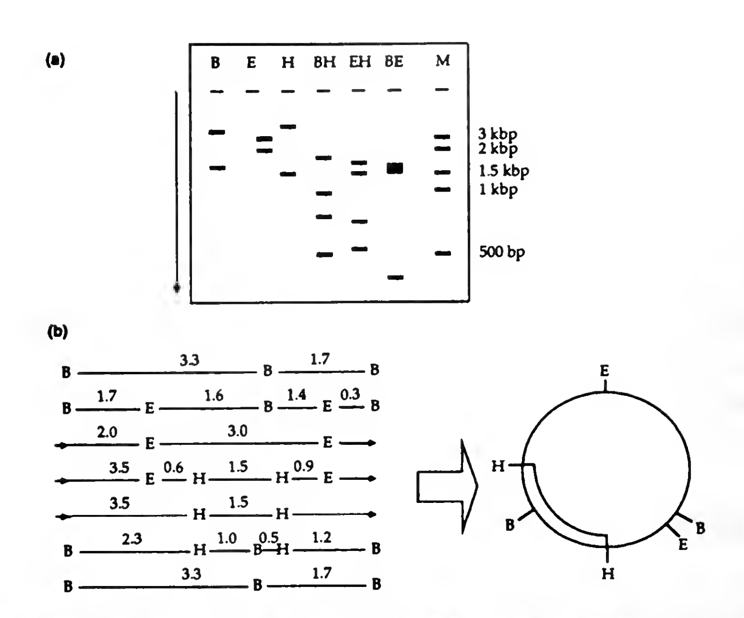


图 24.4 限制性作图的基本原理。通过一系列限制性酶消化载体,根据限制性位点产生物理图谱。通过分析单酶切和双酶切产生的电泳条带的大小(a),能够推导不同位点的相对位置(b)形成完整的质粒图谱。限制性内切酶以它们起始字母表达(B = *Bam*HI, E = *Eco*RI 和 H = *Hin*dIII;BH,BE 和 EH 是双酶切;M 是分子量标准泳道,用已知大小的限制性片段作为对照)。片段大小用 kb 表示。注意当产生类似大小的不同片段,它们在胶上一起迁移,在 *Bam*HI/*Eco*RI 双酶切中就有这种情况。

转录分析　**转录分析**是在 RNA 水平研究基因结构和表达。对基因结构和表达的分析很重要,所使用的方法在表 24.6 中讨论。**Northern 杂交**是检测一个基因在给定组织中是否表达和表达水平如何的有效方法。它也显示转录本的大小和不同的产物(也就是不同的剪接)。基因分析中重要的一步是确定转录起点和终止(或多腺苷酸化)位点以及内含子/外显子的边界。这样精确的结构细节可以通过**核酸酶作图**(nuclease mapping)和**引物延伸**(primer extension)获得,它们是利用 mRNA 和对应基因组 DNA 之间结构差异的技术(图 24.5)。最有效的转录本分析技术之一是原位杂交,它能确定 RNA 在细胞和亚细胞中的表达方式的特点(见文框 24.3)。

表 24.6 在 RNA 水平上分析基因结构和基因表达的技术

方法	原理和在 RNA 结构和表达研究中的作用
Northern 杂交	原理:总 RNA 或 poly(A)$^+$ RNA 电泳分离后转移到滤膜上,与互补的探针杂交 结构:显示转录本大小和是否存在不同的剪切,但只对相对丰富的 RNA 有用 表达:提供了在不同细胞株或组织中 RNA 水平的粗略比较
核酸酶作图 (图 24.4)	原理:是利用剪切单链核酸但不作用双链的酶[来自于稻曲霉(*Aspergillus oryzae*)的核酸酶 S1, RNase A]。将基因组限制性片段标记与 RNA 杂交。双链区域(基因组 DNA 和 RNA 相共线)受到保护,免受核酸酶活性作用,但 RNA 单链中尾巴不受保护 结构:比较核酸处理和未处理样品变性片段大小的区别,能确定转录/非转录序列的边界(也就是转录起点)和内含子/外显子边界 表达:RNase 保护分析也可以作为确定 RNA 水平的敏感方法

续表 24.6

方法	原理和在 RNA 结构和表达研究中的作用
引物延伸 (图 24.5)	原理:引物(寡核苷酸引物或一个克隆的限制性片段)在 mRNA 的 5′端附近退火,通过反转录延伸 结构:对延伸和未延伸的样品变性片段大小差异作比较,确定转录起始点。在 5′异质性和起点选择中尤其有用(参见多起始位点、不同启动子选用)
原位杂交	原理:用标记的 cRNA 探针与内源性 RNA 进行原位杂交,也就是在正常细胞中定位 表达:在细胞和亚细胞水平确定表达模式的有效方法。这种技术可用于组织切片,分离的细胞或更小的样本,全样载片(需要非同位素探针)。对组织切片或分离细胞的原位杂交可用于确定表达水平
RT-PCR	原理:以 mRNA 反转录产生的 cDNA 为模板进行 PCR 结构:反式 PCR(参阅)通过外显子专一性引物的扩增来确定内含子/外显子结构。PCR 最重要的应用是转录本分析,通过 RACE(参见)分离全长的稀有 cDNA 克隆 表达:定量 RT-PCR 可用于确定特定 RNA 分子,即使它们量很少。原位 PCR 确定表达模式。这些技术进一步的讨论见聚合酶链式反应(PCR)

见图 24.5 核酸酶作图和引物延伸分析的一些例子。

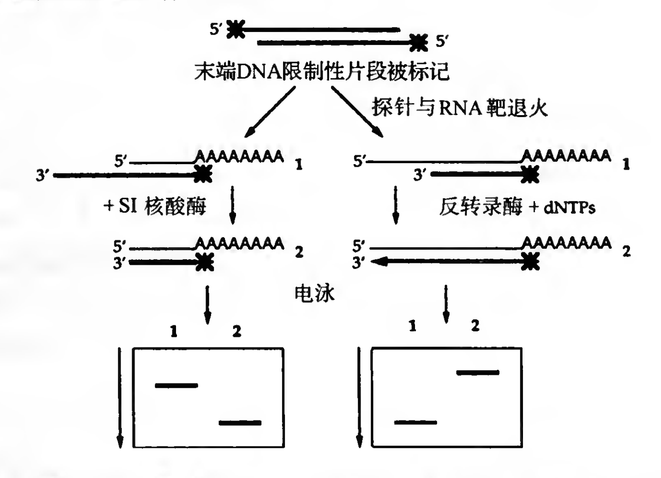


图 24.5 基因转录起始位置的转录本分析。末端标记的限制性片段变性后反义 DNA 链与 RNA 杂交。在核酸酶保护实验中,限制性片段在起点外保持单链的部分被单链专一性核酸酶(如 S1 核酸酶)所降解。在引物延伸实验中,任何限制性片段引物之外的单链 DNA 可以通过引物的反转录复制延伸来复制。在两种情况中都可以比较处理和未处理样品电泳中大小区别大致确定转录起始点。若在相邻泳道上的片段为序列梯(sequencing ladder)[1]就可以得到转录起始精确位置。同样的技术可用于确定 RNA 和 DNA 任何结构的差别,如内含子/外显子边界。DNA 以粗线表示,RNA 以细线表示。

24.4 克隆 DNA 的表达

克隆 DNA 表达的基本原理 上面讨论的载体设计用于 DNA 片段克隆和分离,很多质粒和λ载体有功能性调节元件,使供体 DNA 得到表达。表达克隆的基因有几个目的:①产生大量标记 cRNA 作为探针使用;②构建表达文库(expression libraries)(cDNA 文库中克隆的供体 DNA 是被表达的,能够根据所编码多肽结构或功能的性质可进行筛选);③)在蛋白水平上分析和利用基因的功能;④商业性生产蛋白;⑤制备抗体;⑥捕捉相互作用的分子。

---

1) 序列梯是 DNA 测序凝胶的放射自显影影象。——译者注

在大肠杆菌中进行**表达克隆**较为方便，因为它是主要的克隆宿主，通过基本克隆载体上加上合适的转录和翻译调控顺序使它变为表达载体，有利于供体 DNA 的过度表达。大肠杆菌质粒表达载体带有强启动子，经常是可诱导的，如 *lac* 或 *trp* 操纵子的启动子或 T7 晚期启动子（T7 RNA 聚合酶基因从可诱导的启动子表达）。表达载体还包含了转录终止子和含有保守 Shine-Dalgarno 的顺序核糖体结合位点。这些元件的排列通过优化产生稳定的 RNA 和大量的蛋白。

各种表达策略中的共同问题是表达载体的稳定性。表达大量外源蛋白（或即使是内源蛋白）的细胞比野生型细胞生长缓慢，因此存在对表达载体丢失或表达水平下降突变的选择。可诱导启动子的使用有助于避免这一问题，当大量含载体细胞培养后能够生长，表达可被诱导并很快被收集。表达策略另一重要方面是控制载体分配和设计限制可能自发产生结构突变的表达载体。在载体上加入前导肽（参阅）序列使表达的蛋白从细胞中分泌出去。这可以通过将过度表达的蛋白从细胞内环境中排除而增加稳定性，如果表达的蛋白是有毒性的话，对这对细胞的存活可能是必要的。无论如何，分泌有利于从培养细胞中纯化蛋白。在表达克隆位点的上游含有前导肽的载体称为**分泌载体**（secretion vector）。

天然和融合蛋白　根据特定的应用，克隆基因可以表达产生**天然蛋白**（以其自然状态存在）或**融合蛋白**，融合蛋白是将外源多肽融合到载体编码的多肽一起，如细菌的前导肽（如上文所讨论的）或象 β-半乳糖苷酶这样大蛋白的一个片段。天然蛋白更适合治疗用途，因为融合蛋白可能是对人体有免疫原性。然而融合蛋白在大肠杆菌中经常更稳定，因为它们类似于内源性蛋白，而天然蛋白可能引起降解。融合蛋白还有其他优点：它们易于纯化（如和识别载体携带多肽的抗体的亲和性）和载体多肽的功能在融合产物得以保留，使其可以作为报告分子（reporter）。有时可以用特殊的蛋白酶将载体衍生多肽从融合蛋白上切下，产生天然蛋白。

对于天然蛋白的合成，供体 cDNA 插在表达载体转录和翻译调节元件的下游，并经常携带它自身的起始密码子。对融合蛋白的合成，供体 DNA 的开放阅读框插入时必须与载体基因片段读框一致，这样它们能被连续阅读。这可以通过小心地选择亚克隆的策略来达到，但更方便是用随机方法产生一个正确的阅读框架。合适的策略包括使用三个不同阅读框架克隆位点的载体，使用下游的融合蛋白并筛选对插入片段通读的克隆，或通过将供体 DNA 插入有随机长度同源多聚尾巴的载体中，它可以产生随机长度的同源多聚物的连接。

在大肠杆菌表达克隆的缺点　虽然克隆的基因在大肠杆菌中的过度表达有利于工业规模合成很多原核和真核蛋白，但这种体系还是有很多的问题。如上所述，天然真核蛋白可能在大肠杆菌中不稳定，降低蛋白产生的量，有时达到非常低的水平。另一个问题是过度表达的外源蛋白可能会形成**包涵体**（inclusion bodies），包涵体必须通过强烈的化学处理才能打破，尽管包涵体的大小和数目能通过在较低温度下生长和大肠杆菌分子伴侣（参阅）GroEL 和 GroES 的过度表达得到改善。

即使外源蛋白表达成功，但它可能没有像天然的、内源性合成的蛋白同样的功能。一些过量表达的真核蛋白以活性形式存在（如集落刺激因子），而很多则不是（如表皮生长因子 EGF），反映了缺乏正确的翻译后加工。一旦一个多肽被表达，它必须正确折叠和加工

才具有功能(见第 22 章蛋白质:结构、功能与进化)。然而大肠杆菌经常不能折叠和加工真核蛋白——它不能对蛋白剪切或糖基化,或不能形成正确的二硫键,可能是由于缺乏真核细胞中存在的分子伴侣(参阅)。即使多肽能有效表达和稳定存在,在这些任何步骤中的缺陷都可能使蛋白没有功能。大肠杆菌中表达的活性蛋白可能反映了它们有限的结构限制,例如一些糖蛋白,像白介素-3 无论是否糖基化都有相当的活性。

真核表达宿主 当细菌细胞不能正确地加工表达蛋白时,真核细胞可用作表达的宿主细胞。在真核细胞中表达蛋白也使蛋白分析能在真核环境中进行,使用细胞内的信号序列使蛋白导向于特定的细胞器中(见第 22 章蛋白质:结构、功能与进化)。三种类型的真核宿主细胞用于蛋白表达:酵母、昆虫和哺乳动物细胞,各有与之适应的表达载体和收获策略。

酵母经常是表达那些在细菌中表达效果较差的真核基因的第一个选择,酵母是微生物,遗传背景清楚、易于操作,能用于大规模培养。酵母不仅能表达外源真核蛋白,而且它们提供了研究真核蛋白功能的合适环境[**替代遗传学**(surrogete genetics)],以及研究蛋白-蛋白相互作用的体系(参见酵母双向杂交体系),最后它们提供了用于基因组分析的大容量克隆载体(参见酵母人工染色体)。所以人们对发展酵母克隆和表达载体有特殊的兴趣(文框 24.5)。酵母表达载体经常是基于 2μ 质粒(YEps),尽管 YIps 可用来将 DNA 稳定整合到基因组中。表达载体包含强的组成型或可诱导的酵母转录控制序列(如磷酸甘油酸激酶基因启动子),经常与 2μ 质粒终止子和酵母营养缺陷型(参阅)恢复的筛选标记结合在一起。酵母蛋白很少分泌,但大多数翻译后加工发生在分泌途径中,所以需要用酵母分泌载体来表达糖基化的哺乳动物蛋白(毒性蛋白也是如此)。这样的载体携带酵母 5′前导肽信号序列,这序列是用于分泌交配因子信息素(参见交配类型选择)。

酵母细胞不能提供哺乳动物蛋白功能所需的所有正确翻译后加工。很多哺乳动物蛋白的功能与不同的糖链结构的模式是相关联的,但这些表达产物如果发生不正确的加工,用于治疗经常有免疫原性。酵母糖基化突变体的使用有时可以解决这一问题。**杆状病毒表达系统**(baculovirus expression system)提供了在昆虫细胞培养中,多角体蛋白基因强启动子控制下表达外源基因的途径。在这种体系中产生的蛋白是惊人的——从 10^9 感染细胞中有高达 100μg 的重组蛋白——翻译后加工的程度也得到了很好的证实。杆状病毒系统的其他优点包括它的容量——大到 20kbp 的插入顺序可以被成功地克隆——多个外源基因可以同时表达,能够对蛋白复合物进行表达和分析。

哺乳动物细胞提供了人类治疗蛋白表达的最佳体系,但效率较低。很少哺乳动物表达载体是以基因组外的形式存在,所以长期表达需要构建整合到基因组中的载体(参见瞬时转化、稳定转化)。哺乳动物表达载体的构建是基于细菌质粒,转染细胞的质粒或哺乳动物病毒,它们将 DNA 导入细胞中。大多数病毒载体是重组的整合病毒(如反转录病毒、腺病毒),尽管含疱疹病毒复制起点的载体是基因组外存在的。有效的哺乳动物表达载体携带强的组成型病毒启动子(如 SV40 的早期启动子和增强子)或可诱导的启动子(如热休克启动子或修饰的大肠杆菌 *lac* 启动子)。异源的可诱导启动子的优点在于只有外源基因被诱导,而其他内源性哺乳动物基因不能被诱导。表达载体同时也携带了强翻译起始顺序(参见 kozak 保守序列)、多腺苷酸化位点,经常还有对有些基因有效表达很重要的一个内含子。有些基因通过药物选择能有效维持很高的拷贝数,如二氢叶酸还原酶

(*DHFR*)基因在细胞中能暴露在氨甲蝶呤中(见文框 15.3)。对扩增区域(**扩增子**,amplicons)的分析显示它们比扩增基因本身包含更多的 DNA,也就是说非选择的 DNA 被一起扩增了。含有外源基因克隆位点邻近的 *DHFR* 基因载体被整合进基因组内,并进行选择性扩增。外源基因被与 *DHFR* 一起被扩增,得到很强的表达。

24.5 基因调节的分析

报告基因 克隆基因表达的基本原理是将基因亚克隆在有合适调节元件的载体中,与此类似通过将被克隆的调控元件亚克隆到一个提供可调节合适基因的载体中也可得到表达,有利于分析基因的调节。从原理上讲,任何基因都适用,但使用一个用简单的测定方法就能检测和定量的基因更方便,而且它的产物在宿主细胞中又不存在。**报告基因**(reporter genes)的例子在表 24.7 的目录中列出。不同的报告载体对于分析不同的调节元件是有利的:**启动子探测**(promoter probe)载体含有一个报告基因在多克隆位点中,用于插入和测定假定的启动子元件;**增强子探测**(enhancer probe)载体包含由基本启动子驱动的报告基因和一个插入了假定增强子元件的多克隆位点。也有**终止子探查**(enhancer probe)载体,用于分析细菌转录终止子。

表 24.7 报告基因的比较

报告基因(和产物)	使用
来自大肠杆菌的 *lacZ*(β-半乳糖苷酶)	广泛使用的报告体系。酶将发色底物 X-gal 在基因表达的位置转化成蓝色沉淀。通过 ONPG 转变成可定量可溶性黄色产物。由 IPTG 诱导(参见 *lac* 操纵子)
来自萤火虫的 *luc*(荧光素酶)	当暴露于荧光素底物中时,产生生物荧光的高灵敏报告基因
来自大肠杆菌 Tn9 中的 *cat*(氯霉素乙酰转移酶,CAT)	体外分析中非常有用的报告基因,但在原位分析中分辨率低。CAT 乙酰化氯霉素,乙酰化产物可以通过 **CAT 测定**(CAT assay)中的薄层层析被鉴定
来自大肠杆菌的 *GUS*(β-葡糖醛酸糖苷酶)	一般在植物系统中用的报告基因,酶将发色底物 X-gluc 在基因表达的位置转化成蓝色沉淀
来自水母 *gfp*(绿色荧光蛋白)	一种报告分子,因其自身荧光用在活体系统中有独特的优点

[a]X-gal=5-溴-4 氯-3 吲哚-β-D-半乳糖苷;X-gluc=5-溴-4 氯-3 吲哚-β-D-葡糖苷酸;ONPG=邻硝基苯-β-D-半乳吡喃糖苷;IPTG=异丙基硫代-β-D-半乳糖苷。

基因调节的分析涉及一系列报告结构(有上游克隆调节元件的报告基因)活性的比较,其中调节元件可以通过体外诱变(参阅)进行修饰。这样的分析可以通过用不同的细胞提取物进行体外转录,或将报告基因结构瞬时转染到细胞中或多细胞生物中以及生殖细胞系(**报告转基因**,参见转基因动物和植物)。体外转录和细胞系的方法是简单的实验,所能提供的信息是有限的。细胞系方法的一个特别的缺点是瞬时的基因组外载体不能精确表明体内的调节条件,因缺乏染色质结构和远处顺式调节元件水平的调节。另外,载体的高拷贝数会抵销转录因子(参见压制)。报告转基因生物提供了内源性基因调节更精确的表现度,使调节元件的空间和时间的效应能够显露出来。然而,报告转基因可能有位置和剂量效应,产生额外或限制的表达模式和不同的表达水平。

DNA-蛋白相互作用　比较带有不同突变的报告基因结构有可能非常精确地确定假定的调节元件的位置。然而，为了精确定义这些元件，报告基因结构必须得到如下补充：①顺序分析确定功能性；②证明蛋白-DNA相互作用的实验。有几种体外的技术用来分析蛋白-DNA的结合，包括电泳迁移率变动分析(electrophoretic mobility shift assay)、DNase I足迹法(DNase footprinting)和甲基化干扰(methylation interference)(这些技术的进一步讨论见核酸结合蛋白)。对DNA中蛋白结合基序的鉴定也提供了分离编码新的转录因子克隆的直接途径。对应于假定的调节元件的寡核苷酸能够被标记，并用来在表达文库中筛选相互作用蛋白，这样的方法称为southwestern筛选(见第17章核酸结合蛋白)。

包载载体　报告基因被广泛用于分析克隆的调节元件，但它们还可以用于分析未知内源基因的表达模式。**包载载体**(entrapment vectors)是包含报告基因的结构，它整合到基因组的随机位置，通过产生报告基因的表达模式，对附近的顺式作用调节元件作出反应。有三类包载载体(图24.6)，它们的性质列在表24.8中。这种技术最早被用于果蝇中使用的重组P因子载体，但现在已应用到其他生物，包括植物(*Ac-Ds*因子)和小鼠(重组体反转录病毒或者简单随机整合的DNA；参见随机插入转基因)。包载技术的发展产生了改进的基因包载载体，它可以作用特定类型的基因，如那些编码分泌蛋白的基因。

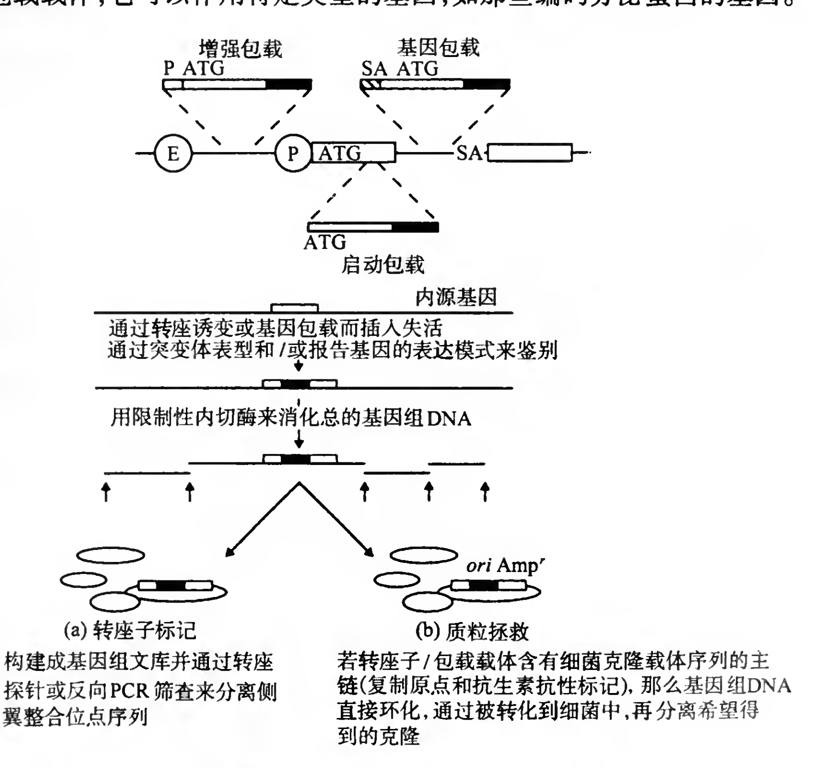


图24.6　包载载体。(a)有三类包载载体：**增强子包载**(enhancer trap)，携带一个基本的启动子，对内源性增强子元件反应；**基因包载**(gene trap)，携带一个剪接受体位点，对内源性剪接供体位点反应使内含子插入；和**启动子包载**(promoter trap)；携带一个简单的起始密码子，被插入的第一个外显子激活。(b)克隆包载基因的策略包括从基因组文库中分离标记的DNA(通过杂交或反向PCR)和质粒拯救。增强子包载载体可能整合在受增强子影响内源性基因的远处，对内源性基因的克隆涉及染色体步移。

表 24.8　包载载体

包载载体	结构	性质
增强子包载	报告基因上游有基本启动子	通过插入内源性增强子的邻近位置激活。方向是独立的,不一定依赖于内源性靶基因的表达。偶尔有突变产生。经常可能通过从基因组文库中载体插入位点的染色体步移来克隆内源性基因(参见*定位克隆*)
基因包载	报告基因上游有剪接受体位点	通过插入内含子激活。取决于周围基因的方向、阅读框架和表达。经常由于外显子插入而突变。基因可以通过基因组或 cDNA 文库的*转座子标记*(参阅)被克隆
启动子包载	报告基因上游有翻译起始位点	通过插入外显子激活。依赖于周围基因的方向、阅读框架和表达。经常由于外显子插入而突变。基因可以通过基因组或 cDNA 文库的*转座子标记*(参阅)被克隆

基因和启动子包载载体产生融合蛋白,并且是阅读框架和方向依赖性的。报告基因上游对*内部核糖体*(参阅)进入位点的应用可以消除这些限制。

使用任何一种包载载体,都有可能直接克隆激活包载的内源性基因,因为它已经被载体顺序所**标记**(tagged)(图 24.6)。转座子标记是分离转座子插入引起基因突变的广泛使用的技术,包载载体(其中很多是重组转座因子)能以同样的方式起作用。载体作为筛选基因组文库的探针,分离包含载体序列和携带基因组 DNA 侧翼插入位点的基因组克隆。这被用于鉴定重叠基因组克隆和同类的 cDNA。如果载体包含质粒克隆载体的骨架序列,简单地消化基因组 DNA 会产生包含两侧基因组 DNA 的质粒。这种技术称为**质粒拯救**,能快速对感兴趣突变和报告基因的表达模式进行分子水平的分析。

24.6　蛋白分析与蛋白间相互作用的分析

用抗体检测和纯化蛋白　　抗体与其对应抗原之间的相互作用是高度灵敏和专一的。与核酸探针 DNA 和 RNA 水平的应用方法相同,抗体可以作为探针在蛋白水平上研究基因表达。原则上,抗体有三种用途:①分离和纯化蛋白(如亲和层析和免疫共沉淀);②在膜上或原位检测蛋白(如 western 杂交,表达文库的免疫化学和免疫学筛选);③干扰或修饰蛋白功能(参见*体内和抗体酶*)。对这些技术的进一步讨论见第 22 章蛋白:结构、功能与进化。

酵母双向杂交体系　　分析蛋白－蛋白相互作用的传统方法包括**免疫共沉淀**(抗体结合后沉淀抗原,洗脱和分析任何相互作用的蛋白)和对抑制基因和增强子突变的筛选(与特定突变相联系的表型分别改善或增强,常鉴定相互作用的蛋白,见文框 15.4)。**酵母双向杂交系统**是基于表达文库体系,用来分析蛋白－蛋白相互作用。它利用真核蛋白基序性质的特点,特别是转录因子的 DNA－结合区和激活结构域可以单独起作用的特点(参见*结构域交换*)。进而几个转录因子被鉴定,它们的 DNA－结合区和激活结构域被分离的基因所编码,如此非共价的结合有利于转录激活(如 Oct－1 和单纯疱疹病毒 VP16 反式激活子;见转录)。酵母双向杂交体系检测与一个已经克隆基因产物相互作用的蛋白,使分离编码这些未知作用蛋白的 cDNA 变得容易(图 24.7)。已知的基因以与经典的 DNA－结合结构域,如 GAL4,形成融合蛋白的形式表达——这是相互作用的诱饵。构建的表达文库中 cDNA 以与转录激活结构域融合蛋白的形式存在,任何**相互作用因子**,即与诱饵相互作用的蛋白,将激活结构域带到 DNA－结合结构域一起。体系的最后成分是报告基因,它被杂交的转录因子激活。相互作用因子这样通过报告基因的表达来鉴定,而对应的

cDNA 可以被分离鉴定。

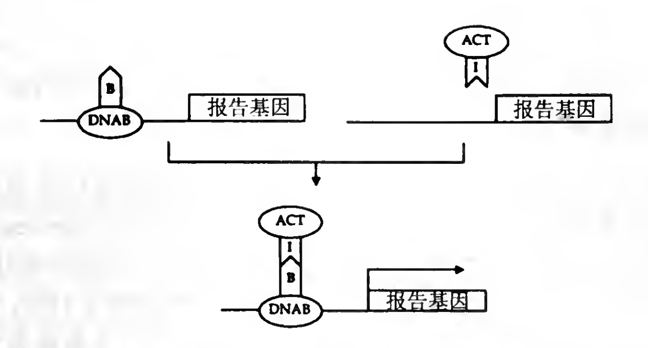


图 24.7 酵母双向杂交体系的原理。实验的目的是鉴定与诱饵蛋白相互作用的蛋白(B)。它以与 DNA-结合结构域(DNAB)的融合蛋白来表达。表达文库的克隆是构建成与转录激活结构域(ACT)的融合蛋白。用文库中的表达载体和包含诱饵蛋白 DNA 结合结构域结合位点的报告结构来转化表达诱饵融合蛋白的细胞。无论是诱饵蛋白或假定的相互作用蛋白(Ⅰ)都不能单独激活报告基因。然而任何与诱饵蛋白相互作用并能激活报告基因的转录的文库蛋白,可被克隆而得到鉴定。

噬菌体展示　　通过将寡核苷酸克隆到噬菌体衣壳蛋白基因的读框中,使外源多肽表达在噬菌体的表面的技术。产生的**融合噬菌体**保留了它们的感染性,能以正常方式形成噬菌斑。**噬菌体展示文库**(phage display library)像常规表达文库一样可用抗体或其他蛋白作为探针进行筛选。噬菌体展示技术的主要用途是筛选通过随机突变产生的多肽,改进其亲和力或结合的专一性。这在蛋白质工程中有用,如改进有重要商业价值的酶的性能等,和构建重组抗体。在后面一种情况,噬菌体展示等同于人工亲和力成熟(参见*体细胞过度突变*),避免对杂交瘤细胞株的依赖。相互作用的噬菌体颗粒可以从高达 10^8 的非相互作用噬菌体的背景中纯化得到。

24.7　体外诱变

改变克隆的基因　　一旦一个 DNA 分子被克隆,**体外诱变**技术能被用于导入顺序的改变。这可以是:①比较野生型克隆和突变体功能的特定突变,也就是鉴定重要氨基酸残基或调节元件;②在一个特定区域随机突变,筛选变异体,也就是说鉴定那些性能改善的突变(参见*噬菌体展示*)。为了在体内检测功能,体外诱变的 DNA 可用来替代细胞基因组中的或整个动物或植物中的同源序列(参见*置换*、*基因寻靶*)。这是经典遗传分析的相反过程,它在随机突变中产生和筛选突变来选择感兴趣的基因或系统(参见*遗传筛选*)。

导入特定的点诱变　　在克隆 DNA 分子中特定位点导入特异碱基来替换或产生小的缺失称为**定点诱变**(site-derected mutagenesis),有多种应用途径。**盒式诱变**(cassette mutagenesis)是将供体 DNA 片段用限制性内切酶剪切,并用含所需突变的合成寡核苷酸来替代它。另一种**寡核苷酸诱变**(oligonucleotide mutagenesis)需要将供体 DNA 克隆在噬菌粒载体中,以产生单链 DNA。包含所需突变的寡核苷酸与单链载体退火,产生一个不配对位置。寡核苷酸起 DNA 合成引物的作用,所以载体的*复制形式*(参阅)是异源双链,含一条

有突变的链和一条野生型链[PCR 诱变(参阅)也是应用类似的引物不配对策略]。接下来的复制产生同源双链的野生型和突变载体,后者通过杂交分析被鉴定。这是产生突变的有效机制,但需要关闭宿主细胞的错配修复系统和在野生型中筛选出突变载体。很多**链选择策略**(strand selection strategies)使这一过程更为容易,如用抗限制性酶切的硫代核苷酸保护突变链,使野生型链被降解。

系统化突变分析 例如 DNA 中调节序列的功能分析往往从大的缺失和取代突变开始(用限制性内切酶或外切酶),来确定重要区域的位置。**缺失诱变**(deletion mutagenesis)涉及用限制性酶从克隆中切除供体 DNA 的一部分。**单向缺失**(unidirectional deletions)能通过象大肠杆菌外切酶 VII 这样的外切酶产生,这样的外切酶作用于特定类型 DNA 末端的底物。单向缺失可以产生一系列的**嵌套突变**(nested set of deletions)(一个末端是共同的,另一个是可变的)。这在分析调节元件和对点突变作图是有用的。**筛选突变**(scanning mutagenesis)是系统化地对克隆的每一部分替代,来决定它的功能。在分析调节元件时,**接头筛选诱变**(linker scanning mutagenesis)可以进行小片段的缺失,延着克隆在每个位置被寡核苷酸接头替代,保留了剩余 DNA 基序的空间关系。在蛋白功能分析中,**同源筛选突变**(homologscanning mutagenesis)涉及将蛋白的每个片段都用相关蛋白的同源区域替代,来鉴定功能特异性的残基。这一技术可以延用到通过定点诱变系统地将不同的残基替代每个氨基酸(参见结构域交换)。

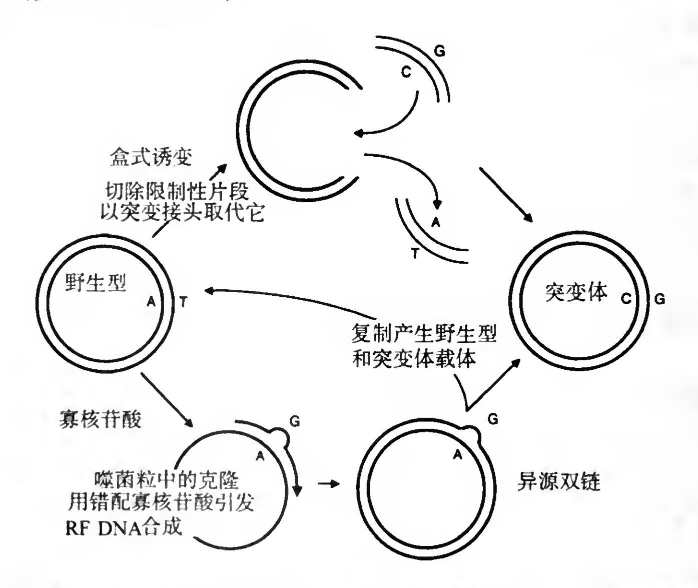


图 24.8 定点诱变。为了在体外导入特定的点突变,小的限制性片段可以切除,并被包含所需突变的合成接头所替代(盒式诱变)。如果这不能办到,插入片段可以克隆到产生单链的噬菌粒载体中,由错配引物指导复制型的合成。这样产生一个异源双链的复制型,在下面的复制循环中产生野生型和含突变的复制型。

随机诱变 定点诱变能精确地导入一个特定的点突变,**随机诱变**(random mutagenesis)是分析大量不同突变体效应的最快速的方法。随机点突变适于不精确的条件下和在所有 DNA 修复系统被关闭的宿主细胞中,可通过使用有倾向差错的 DNA 聚合酶(如反转录酶,*Taq* DNA 聚合酶)来导入。随机突变也可以用简并(参阅)寡核苷酸,以盒式诱变、寡

核苷酸诱变或 PCR 诱变，定向于特定区域。最近通过体外重组对基因随机片段，以 PCR 介导修复的技术已经被发展起来。尽管还未被广泛应用，但这种技术已经成功地产生了活性大大高于野生型酶的β-内酰胺酶的变异体。

24.8 转基因：基因转移到动物和植物

基因转移到高等真核生物细胞的机制　分子克隆最重要的基本技术之一就是将 DNA 转导进入细胞。如上文所讨论的，有几种常规使用的高效方法可以将 DNA 导入细菌和酵母细胞，促进了动物和植物基因的克隆和表达。然而，对动物和植物 DNA 的功能分析经常需要将克隆的 DNA 重新导入原先的物种中。

有一系列技术可以将 DNA 导入真核细胞(表 24.9)，大多数涉及强迫细胞吸收裸露的 DNA(转染)，但有些基因转移技术是基于包装在病毒衣壳中 DNA 的转导，而在植物中有一种独特的基于细菌质粒接合转移的方法(参见第 10 章细菌中的基因转移)。

表 24.9　DNA 导入真核细胞的方法

方法	评论
细胞转染方法(吸收裸露的 DNA)	
聚乙二醇	原生质体融合或在 Ca^{2+} 离子和聚乙二醇存在的情况下直接吸收 DNA。在酵母和植物中，需要从原生质球/原生质体中重新产生细胞，有效但费力
化学转染	有很多方法，如酵母的乙酸锂转染，动物细胞的磷酸钙或 DEAE-糖苷转染。DNA 通过内吞吸收。这些方法一般对瞬时和稳定转染都有效，除了 DEAE-糖苷转染动物细胞，对稳定转染效率较低
脂质体转染法	DNA 与阳离子脂质体形成的复合物被内吞吸收。此是转染动物细胞、酵母和植物原生质球/原生质体的高效方法。经常在其他方法不能起作用的哺乳动物细胞中能发挥作用
电穿孔	通过高电压短暂冲击产生瞬时小孔吸收裸露的 DNA。是转染酵母、植物和动物细胞非常有效的方法
直接注射	是将 DNA 导入细胞十分费力的方法但有 100%效率。通常用于动物卵母细胞、卵细胞和合子，以及其他方法转移困难的细胞
微弹(生物弹)	使用**显微注射**(microprojectiles)，以钨或金的颗粒包被 DNA，用**基因枪**(gene gun)高速打入细胞(原先是改进的短枪，但近来发展了用空气或放电的高压冲击波更精制的装置)。不需要去除细胞壁，能对植物细胞有效转染。也能用于转染整个植物和动物组织
转导方法(DNA 在病毒衣壳中转移)	
重组病毒	主要用于哺乳动物的病毒载体。几种病毒能被用作载体——疱疹病毒在基因组外存在，反转录病毒和腺病毒整合到基因组(参见*杆状病毒表达系统*)。植物病毒也被发展成载体，但由于 Ti 载体的成功，所以植物病毒没有被广泛应用。病毒启动子经常被用在表达载体中，驱动结构基因的大量表达
接合方法(DNA 通过细菌接合转移到真核细胞中)	
Ti 载体	用根癌农杆菌的 Ti 质粒切得的 T-DNA 转化培养的活体植物和植物细胞。重组的 T-DNA 是高效的基因转移载体，尽管只有双子叶植物能被转化(文框 24.6)

转染(transfection)：当用于真核细胞时，任何裸露 DNA 的吸收都称为转染(不是噬菌体 DNA 在细菌中的情况)。相反，**转化**(transformation)是指通过 DNA 整合到基因组中改变基因型(在细菌中，转化是将裸露的非病毒 DNA 导入细胞的过程，与真核生物转染类似)。转染和转化在酵母中的使用是类似的。

大多数细菌和酵母载体是游离在细胞质中，也就是作为自身复制子在基因组外维持。相反，高等真核生物缺乏核质粒和潜在的游离在细胞质中的 DNA 病毒(哺乳动物中疱疹病

毒为基础的载体是例外)。这样的DNA在高等真核生物中不能像细菌和酵母那样以基因组外的形式存在。在有些情况下,这并不重要,因为DNA不需要被稳定地维持:很多实验,如报告基因分析在相对快速和**瞬时转染**中就能进行(当DNA导入细胞后逐渐因稀释和降解所减少)。如蛋白过度表达需要**稳定转染**(stable transfection),在没有游离载体的情况下,一般通过DNA稳定整合到基因组中达到目的。DNA直接转染高等真核细胞时频繁地随机整合到基因组中,提供了在培养的动物和植物细胞中遗传转化的相对简单的机制。

转基因动物和植物　在多细胞生物中,将上面所讨论的基因转移技术应用到**全能**(totipotent,参阅)细胞中,能够产生体细胞基因型有相同改变的动物和植物。这样的生物被称为**转基因的**[①] (transgenic),可通过生殖系统以简单的孟德尔特征传递它们新获得的遗传性状。

转基因植物的路线相对简单,因为Ti质粒是自然发生的其主要基因转移系统是高效的。还因为分化的植物细胞是全能性的,至少在有些物种中,能够在合适的条件下重新产生整个成熟的植物(文框24.6)。相反分化的动物细胞的全能性有限制,因此转基因动物必须通过对卵细胞或从早期胚胎得到的细胞进行操作才能产生。有三种方法可以产生转基因动物(文框24.7)。使用最广泛的是直接将DNA注射到卵细胞的细胞核中,导致DNA随机整合到基因组中。在小鼠中,转基因技术更精细,胚胎干细胞(ES)的**基因寻靶**(gene targeting,参阅)能够进行特殊的遗传改造。与胚系转化一样,基因转移到活体体细胞的技术也被开发出来。**体细胞转基因**(somatic transgenesis)最雄心勃勃的应用是**基因治疗**(gene therapy),是用核酸治疗甚至纠正人类疾病(文框24.8)。

转基因动物和植物提供了研究基因功能的难以比拟材料。与类似的体外或在培养细胞中进行的研究不同,在转基因生物插入的遗传物质,**转基因**(transgene)能在整个生物体不同阶段被研究(即不同的空间和发育阶段的表达)。这可以用于分析缺失和获得基因所产生的功能效应并可检测调节元件。尤其重要的是可以用转基因小鼠作为人类疾病和肿瘤的模型。另一个转基因技术的重要用途是大量表达外源蛋白。**动物产药**(animal pharming)是表述在动物的乳液中生产有商业价值蛋白的委婉用词——特别是药物——通过内源性乳液蛋白基因启动子驱动的转基因。这种方法提供了一个丰富和可再生的重组蛋白来源,这种重组蛋白是以易于纯化的状态存在。类似的转基因植物提供了生产化学制品、燃料、药物和新型食品的来源。转基因技术还能通过加进新的特征增加商业上重要的动植物的性状(如对除草剂和害虫的抗性)和对已存在的性状能进行改进(如水果、谷物、肉类和牛奶的产量和质量)。

转移到真核细胞中的DNA的命运　如上文所讨论的,转移到真核细胞的DNA很少游离存在于细胞质中,而是整合到基因组内。外源性DNA能以三种方式与基因组DNA相互作用:①非法地重组随机整合;②它可以通过单个交换的同源重组整合到一特定位点(共整合或融合);③它可以通过双交换同源重组或基因转变来替代基因组的一个片段(**置换**)。

---

① 转基因以前定义为外源基因(即来自于不同物质的基因)导入一个动物或植物的胚系中,而现在的定义已变成涵盖了各种形式的胚系操作,包括一个内源基因额外拷贝的导入,定向破坏及反义基因的导入。有些研究者用定向这个术语而不用转基因来描叙动物和植物,即使它们的胚系差异是微小的(以双重取代的策略来产生点突变),或者即使是有关基因来源于同一物种。

在酵母中,同源重组尤其有效,大多数基因组的整合是单交换或置换。在哺乳动物中,由于高活性的末端连接修复系统的存在,非法重组比同源重组发生频率高 10^5 倍。因此,大多数导入动物和植物细胞的 DNA 是随机整合在偶尔发生的 DNA 链断口位点,对很少的同源重组需要仔细选择策略(图 24.9)。

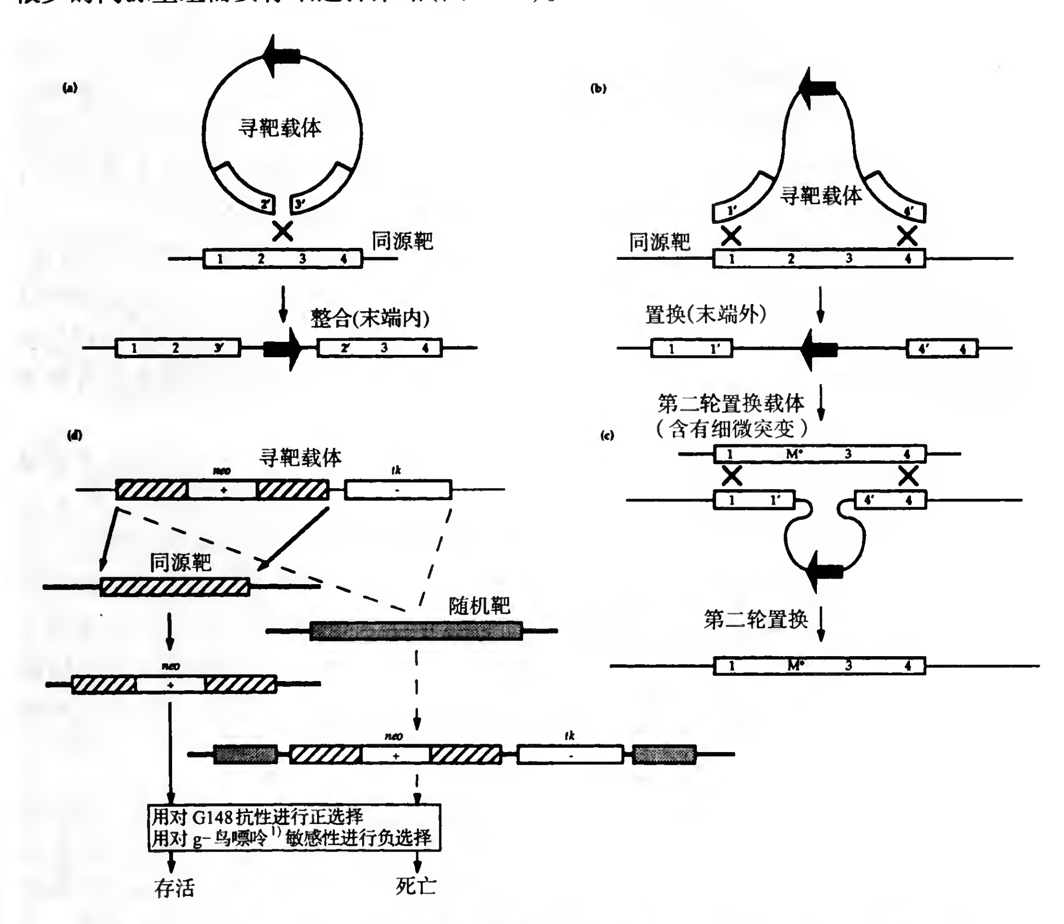


图 24.9 基因寻靶。定向 DNA 的整合可以通过两类剔除载体中的一种:(a)插入载体(靶的内部带有末端单交换位点)或(b)置换载体(靶的外侧带有末端双交换位点)。插入载体完全整合到基因组中,而置换载体用同源的载体序列替代基因组的部分序列。在两种情况中,载体的大片段留在基因组中,因为需要使用显性选择标记(箭头表示)。为了完成精细的定向诱变,如点突变(以 M * 表示),需要第二轮替代(c)。在哺乳动物基因组中,同源重组很少发生,而随机整合很平常。使用双正 - 负选择(d),如用大肠杆菌的 *neo* 基因和疱疹病毒 *tk* 基因。正选择的 *neo* 基因对抗生素 G418 有抗性,而 *tk* 基因对胸苷类似物 9-[1,3-二羟-2-丙氧甲基]鸟嘌呤[1]敏感。*tk* 基因位于寻靶载体的同源结构域侧外,所以只能通过随机插入导入基因组中。只有那些进行同源重组的细胞才能对 9-[1,3-二羟-2-丙氧甲基]鸟嘌呤[1]和 G418 有抗性。

专门设计用来将 DNA 导入基因组的载体称为**基因送递载体**(gene delivery vector)或**自杀载体**(suicide vector)(后者的意思是因为它既不够维持也不能恢复)。如果宿主基因组和载体有同源,相互之间会有同源相互作用,这样的相互作用被载体的线性化所促进,

1) 9-[1,3-二羟-2-丙氧甲基]鸟嘌呤(ganciclovir,GCV)是无环鸟苷(ACV)的类似物,但在抗巨细胞病毒方面更为有效。GCV-三磷酸通过抑制病毒 DNA 聚合酶而抑制病毒 DNA 的合成。原文写成 gancylovir。——译者注

这是因为线性化的 DNA 通过侵入同源双链启动重组(见第 25 章重组)。

随机整合转基因——获得功能效应 诱导动物或植物基因组改变的最简单策略是将外源 DNA 随机整合。这对大多数生物系统是唯一的有效手段,且足以研究或得到那些转基因整合产生的功能获得和显性功能丢失两效应。

获得功能的效应不仅是由于在基因组中导入一个外源基因,还可以增加了一个内源性基因的表达水平或范围(如通过导入在强病毒启动子控制下的额外拷贝——**过度表达研究**——或不同的细胞类型专一性启动子——**异位表达研究**)。外源基因的表达经常是生物技术工程的基本目的,然而研究机构倾向于分析基因功能和调节。对在正常或经过修饰的调节元件控制下报告基因表达的分析在研究基因调节中被广泛应用。

转基因的随机整合方法有几个相关问题:①转基因经常有位置效应,可能引起沉默、改变异位表达或限制表达模式和改变表达水平;②转基因有剂量效应——不能控制整合的拷贝数;③ 在哺乳动物中,特别是在植物中,在基因组中整合了多拷贝 DNA 序列会引起外遗传沉默现象(见 DNA 甲基化和外遗传调节);④可能对内源性基因的表达有非专一性的效应。

最近转基因技术的改进有助于减弱整合的位置效应。这些反映了:①整合位点的异源调节元件和染色质功能域结构的影响;②转基因通常较小且缺乏远距离调节元件,在正常情况下使这些位置不依赖调节元件。在动物和植物中发现,通过在转基因两侧加上边界元件(参阅)可以降低位置效应,可能是由于将转基因特定作为独立的染色质结构域所致。最近发展的 **YAC 转基因**,小鼠携带酵母人工染色体转基因,使大片段的 DNA 能进入小鼠基因组,这样,基因有机会受到各种内在调节元件的影响。YAC 转基因对大基因和大范围调节现象的研究有极大的价值,如基因座控制区域和增强子的活性、染色质结构域效应,亲本印迹和体细胞的超变。

随机整合转基因——失去功能效应 对失去功能效应的研究经常需要定向破坏特定的基因,接着繁殖产生纯合体(见下一部分)。然而随机整合转基因也能用于基因功能失去的研究,尽管经常只在是对野生型显性的情况下(因为导入一个隐性突变等位基因到基因组不会有效果)。偶尔一个随机整合的转基因会碰巧打断一个内源性基因(**插入失活**),在这些情况下杂合体[显性突变经常是由于单倍体机能不全(naploinsufficiency)]或纯合体(隐性突变)中可产生表型。这是诱变的初步和偶然的形式,是不定向的。然而转基因整合的随机插入诱变原理可用于大规模的遗传筛选(参见*转座子标记,基因包载*)。

显性功能丢失效应能够通过多种途径产生:①如果一个突变等位基因以显性失活方式作用,一个随机整合转基因会破坏野生型等位基因的功能;②选择性**细胞部分切除**(cell ablation)可以通过表达一个在组织特异性启动子控制下的毒蛋白,象蓖麻毒素;这可用于研究表达特定基因的各种细胞杀伤效应;③显性或部分显性的**基因击倒**(gene knock-down)效应可以通过表达反义 RNA 或核酶结构寻靶于一个特定基因来达到——这可以通过降解或灭活 mRNA 来抑制基因功能(文框 24.8);④类似的是在蛋白水平的基因击倒可以通过表达一个重组抗体来达到,结合并抑制特定蛋白的活性(见文框 24.8)。

通过同源重组进行基因寻靶 **基因寻靶**(gene targeting)是体内定点诱变的一种形式,涉及到含有一个等位基因的**寻靶载体**(targeting vector)和具有不同等位基因内源基因间

的同源重组。有两类寻靶载体供使用：**整合载体(末端-内侧载体**，ends-in vectors)，在同源结构域中剪切，促进产生一次单交换，使整个载体整合；**置换载体(末端-外侧载体**，ends-out vectors)在同源结构域外侧线性化和双交换或在同源结构域内发生基因转换以载体的同源区域替代基因组的一部分(图 24.9)。

基因寻靶有多种应用：

(1)**基因剔除**(定向破坏)可以通过在整合载体中任何位置或在置换载体的同源结构域中插入一个 DNA 盒来完成(在图 24.9 中黑色箭头表示)。这个 DNA 盒经常是一个显性选择标记，如细菌的 *neo* 基因，能够用来选择靶细胞。

(2)**等位基因替代**。一个等位基因被另一个替代，如用来研究精细突变的效应。这需要两轮替代，因为选择的需要意味着整合载体和置换载体在基因组中留下载体顺序(图 24.9)。

(3)**基因打入**(gene knock-in)是用一个基因替代另一个基因的新应用。这通过将进入基因作为盒子插入同源结构域来实现，这在多基因家族中不同成员间的交换非常容易。

(4)**基因治疗**。在这种情况中，突变的无效等位基因被正常等位基因所替代(见文框 24.8)。

基因寻靶在酵母中是一个有效的方法，正被积极用于系统化剔除所有 6300 个基因的计划。在小鼠中，基因寻靶是通过转染 ES 细胞进行，比随机整合效率低。正负策略需要选出很少的目标细胞，图 24.9 中显示。尽管这些限制，这种技术在分析基因功能，包括很多在发育过程中起重要作用的基因上还是非常有价值，然而从这种实验中得到了意想不到的发现，发育基因高水平的遗传丰余结果是很多小鼠的无效突变体显示出令人惊奇的中性表型(参见丰余)。

可诱导的转基因活性　通过将转基因生物置于可诱导控制下，可将调控的额外水平设计引入转基因生物中。通常有两种形式的调控被使用：①转换基因表达开或关的可诱导启动子；和②可诱导的位置专一性重组体系，不仅有利于控制基因表达，而且也可以用干细胞类型专一性基因缺失和染色体重排。

可诱导的转基因广泛用于过度表达和异位表达的研究。热休克诱导经常用在果蝇和植物中。在小鼠中很多不同的体系经试验得到不同的结果。异源调节体系最成功，因为很少残留活性并且诱导是专一性针对转基因而不会同时激活内源性基因。例子中包括果蝇的蜕皮激素启动子，它对果蝇蜕皮激素反应，而 **Tet 体系**对四环素诱导作出反应。

位点专一性重组(参阅)是涉及短的保守序列(重组子)和识别并催化它们间重组的酶(重组酶)的重组形式。成对重组子元件特定排列可以促进缺失、倒位和易位(共整合)发生(见文框 25.4)。如果重组酶基因和由编码酶识别的重组子元件被插入转基因生物中，会发生 DNA 定向重排。定向缺失可用于基因剔除(如通过整个基因缺失)或基因重新激活(通过缺失由启动子分离的插入基因)。定向染色体重排也可能产生。这种技术的能力来源于对重组酶的控制。重组酶基因可以细胞类型专一性方式或可诱导控制被激活。在第一种方式中，可以产生细胞类型专一性的基因剔除；而在第二种类型中，可以在生物生命周期的任何阶段产生基因剔除，即若基因被剔除则有多效效应(参阅)但是胚胎致死时非常有用。Cre-*lox* 重组酶被广泛使用，特别是在转基因小鼠中，酿酒酵母的 2μ 质粒 FLP-*FRP* 系统在果蝇中得到良好的发展。这些系统内在功能在文框 25.4 中讨论。

文框 24.1 关键工具和技术Ⅰ:限制性内切酶

酶类型	特征
Ⅰ型	三亚基复合物,分别有不同的识别、内切酶和甲基化酶活性
	Mg^{2+},ATP 和 S-腺苷甲硫氨酸(SAM)为活性所需
	识别位点是二部分组成的,剪切发生在随机位点>1kb 远
Ⅱ型	内切酶和甲基化酶相分离,单亚基酶识别相同目标顺序
	活性需要 Mg^{2+}
	识别位点经常是二重对称的——根据识别位点结构有多种亚型。剪切发生在两条链识别位点的附近
Ⅲ型	在双亚基复合物中内切酶和甲基化酶分开,其中一个亚基是相同的
	活性需要 Mg^{2+} 和 ATP,SAM 刺激并不关键
	识别位点是单个的。剪切位点是可变的,大约在识别位点下游 25bp。剪切只在一条链发生。

限制性内切酶的类型 **限制性内切酶**(**限制性酶**)是细菌识别特殊核酸序列(**限制性位点**)的内切酶,经典的识别位点是长 4~8 碱基对。限制性酶的生理作用是*宿主控制的限制和修饰*(参阅),因此每个限制性酶都与相同的 DNA 甲基化酶相结合来保护宿主 DNA 免受**自我限制**。至少有三类限制性酶(见上表),但只有Ⅱ型酶对构建重组 DNA 分子有作用:它们经常精确地在限制性位点相关的磷酸二酯键剪切 DNA 产生确定的产物——限制性片段。

命名法 限制性内切酶由物种三字母以斜体命名(如 *Eco* = *E. coli*, *Hin* = *H. influenzae*),如果需要,限制性表型是由质粒或噬菌体赋予的,用更多的字母或罗马数字表明株系或载体(如 *Eco*RI,*Hin*d,*Bam*H)。最后,如果同一细胞中多于一个限制性体系存在,它们就用罗马字母加以区别,如 *Hin*dⅢ。如有必要,一个限制性修饰系统的内切酶和对应甲基化酶可以通过前缀 R 和 M 分别定义,如 R.*Bam*HI,M.*BamH*I。

Ⅱ型限制性位点的分布和剪切频率 Ⅱ型限制性内切酶剪切 DNA 的频率是依赖于限制性位点的大小(酶可以被称为 4-切割,6-切割等)。在随机 DNA 序列任何基序中的频率是 $1/4^n$,n 是基序的大小。因此 4-切割酶如 *Sau*3AI(GATC)倾向于每大约 250bp 中剪切 DNA 一次,而 6-切割的像 *Eco*RI(GAATTC)产生平均 4kbp 大小的片段,如 *Not*I(GCGGCCGC)的 8-切割则平均 65kbp 产生片段。片段的大小也依赖于底物的碱基组成。**稀有切割**有大的识别位点或识别位点只在特定基因组中出现。*Not*I 是 8-切割,限制性位点是 GC 丰富的(较少出现在哺乳动物基因组中——40%GC),包含两个 CpG 基序(在哺乳动物 DNA 中是很缺乏的)。因此 *Not*I 消化哺乳动物 DNA 预计的平均片段大小是大约 95kbp。稀有切割在制备粘粒文库和大范围限制性图谱有用(参见*内含子-编码内切酶、HO 内切酶*)。

Ⅱ型限制性位点的属性和限制性片段 Ⅱ型限制性位点是二重对称的。如果剪切发生在对称轴上产生则平末端。如果剪切位置不直接彼此相对,则产生一个交错的断口,是 5′或 3′突出末端(**粘性末端**)。一般情况下相同酶产生的限制性片段是匹配的而不同酶产生的不匹配。一些例外在下面讨论。

相同的限制性内切酶不总是产生匹配的片段。限制性位点可以是专一性的,其剪切核苷酸顺序是不变的,产生匹配的末端(如 *Hin*dⅢ经常在顺序 AAGCTT 剪切)。其他位点包括一个或多个

双关的核苷酸,增加了随机 DNA 顺序中的剪切频率,但意味着酶产生的末端并不总是匹配的(如 *Hin*dⅡ在顺序 GTYRAC 剪切,产生四种不同类型的粘性末端)。限制位点是单个的,限制序列也许是连续的或是二重的,也许显示不完善两侧对称(如 *Eco*NI 在顺序 CCTNNNNNAGG 剪切,N 可以是任何核苷酸)。由于中心核苷酸的随意特性,在双向位置上剪切不产生广泛匹配的片段。在次理想条件下,一些限制性内切酶的专一性会被降低,只有部分正常识别位置被识别。此称为**星状活性**(star activity)(如在次理想的 pH 下,通常识别 GAATTC 位置的 *Eco*RI 只识别内部的 AATT 顺序)。酶 *Bcg*I 是一个独特的酶,它在 DNA 的每条链剪切两次,产生包含识别位点的小片段。而 *Taq*II 是唯一识别两个不相关位点的酶。

不同的限制性内切酶可以产生匹配的片段。识别不同位点的限制性内切酶有时可以产生匹配的粘性末端。这可以发生在一个酶的识别位点包含在另一个大的位点中,是一个**嵌套式位点**(nested site)。*Bam*HI 识别顺序 GGATCC 和 *Sau*3AI 识别内部的四核苷酸 GATC;二者都产生匹配的 GATC 的 5′突出。连接后产生可以被原始酶的一个或两个识别的**杂种位点**(hybrid site),或者两个酶都不能识别(在所举例子中,*Sau*3AI 剪切 *Bam*HI/*Sau*3AI 杂种位点,但 *Bam*HI 的剪切依赖于两侧核苷酸)。不同来源的限制性酶也可以识别同样的位点。这样的酶称为剪切相同位置的**同切点酶**(isoschizomers)和剪切不同位置的**新切点酶**(neoscizomers)或**异切点酶**(heteroscizomers)。*Sma*I 和 *Xma*I 都识别六碱基位点 CCCGGG。然而 *Sma*I 在对称轴剪切产生平端而 *Xma*I 在第一和第二个胞嘧啶间剪切产生 CCGG 的 5′突出。

甲基化敏感　每个限制性内切酶都有**相关的甲基化酶**(cognate methylase),在宿主基因组中通过甲基化修饰限制性位点,阻止自我限制(参见宿主限制和修饰)。这样所有的限制性内切酶都有一定程度的甲基化敏感性。一些限制性酶由于它们识别位点的特性,对基因组范围的甲基化敏感,如大肠杆菌基因组的 Dam 和 Dcm 的甲基化,和真核生物基因组中 CG 或 CNG 基序的甲基化(见第 7 章 DNA 甲基化及外遗传调节)。甲基化敏感性不同的内切酶对甲基化 DNA 作图非常有用。如 *Hpa*II 和 *Msp*I 识别 CCGG 顺序,但只有前者对内部胞嘧啶的甲基化敏感。这些酶可用于确定高等真核生物中甲基化的 CpG 基序位置(参见 HTF 岛)。

文框 24.2　关键工具和技术 II:凝胶电泳

凝胶电泳　**电泳**(electrophoresis)是根据分子的电荷和大小在电场中分离分子方法。**凝胶电泳**(gel electrophoresis)是分离大分子混合物的标准方法(如蛋白质和核酸),因为在凝胶中没有对流,单独的片段形成清晰确定的条带。样品上样在凝胶一端狭窄区域,这是在配置凝胶时形成的孔。电场穿过凝胶,样品根据大小和电荷以不同的速度从孔中移出。因为所有的核苷酸在磷酸骨架上有同样的负电荷,它们的迁移率只取决于分子大小和形状。蛋白有不同的电荷,根据带电和大小分离,但在去垢剂十二烷基磺酸钠存在下变性蛋白电荷一致化,按照分子量大小分离(参见 western 杂交)。

用于核酸的标准凝胶电泳　DNA 和 RNA 分子从寡核苷酸到 20kbp 的限制性片段可以在标准电泳胶上分离。使用的电泳方法有两类:水平的**琼脂糖凝胶**有中等分辨率,用于分析和制备 100bp 到 20kbp 大小片段,而垂直**聚丙烯酰胺凝胶**具有**单核苷酸**的分辨率,用于分析和制备小分子(如可满足 DNA 测序的需要)。在两种情况下凝胶的平均孔径大小可以通过改变浓度改变,并且不同浓度

用来分离不同大小范围的核酸。在琼脂糖凝胶上的核酸经常用内掺染料溴乙啶染色检测，它在UV 光下发荧光。聚丙烯酰胺凝胶上的条带经常用放射自显影，尽管也可以使用银染。

适用于大的 DNA 分子　　核酸在凝胶中移动时改变它们的构象，在伸展和压缩形式之间变化。它们的速度依赖于凝胶孔径大小和核酸紧密形式球状大小之间的关系，大的分子移动更慢一些。一旦到达临界大小，紧密分子太大不能通过任何孔径，只能以延伸的大分子移动，这样的过程称为**爬行**(reptation)。在这点上，DNA 的移动率不依赖于大小，使所有大分子一起迁移。为了分离如 YACs 和长的限制性片段 DNA 分子，电泳以脉冲电场进行。短暂的电场引起 DNA 分子重新改变方向；长分子比短分子重新排列需要更多时间，这样延迟了它们通过凝胶的时间，使它们能够分离。达到 200Mbp 的 DNA 分子已经被不同的以脉冲电场为基础的方法所分离(在下文中总结)。

2-D 电泳　　电泳在两个方向进行，利用在每个方向分子的不同属性，得到更好的分辨率。2-D 蛋白电泳涉及第一相的等电聚焦(根据在 pH 梯度电荷电分离)，接着加上 SDS 在第二相上根据分子量分离。DNA 的 2-D 电泳使相同分子量但不同构象的分子能够分离(如拓扑异构体、构象异构体、复制中间物)。2-D 电泳涉及第一相根据分子量分离，接着加入溴乙啶诱导构象变化，在第二相中分离结构异构体。

方法(GE＝凝胶电泳)	简单表述
恒定电场方向方法	
脉冲电场(PFGE)	应用短脉冲电场，分子分辨率＜400kbp
电场反转(FIGE)	脉冲和变极性的电场，分辨率＜800kbp
电场方向变化的方法	
脉冲电场梯度(PFGGE)	脉冲和直角交变的电场，分辨率＜2Mbp
直角交变电场(OFAGE)	与上同，改变角度是 45°而不是 90°，改进了条带迁移的表现
反向交变电场(TAFGE)	与 PFGGE 一样，角度变化是整个凝胶反转
钳位均匀电场(CHEF)和程序自动化控制电极(PACE)	凝胶以多边形方式被多个电极包围，分辨率＜7Mbp

文框 24.3　关键工具和技术Ⅲ：核酸杂交

核酸杂交　　单链核酸间互补碱基对成为一些最重要的生物过程的基础：复制、转录、蛋白合成和调节、RNA 剪接、重组和 DNA 修复。**核酸杂交**(nucleic acid hybridization)是描述一系列技术，它们利用双链核酸进行**变性**(denaturation)或**熔解**(melting)(分离成单链)和互补单链自发**退火**(anneal)形成双链的能力。双链 DNA 可以被变性，并且同样的链可以**重新退火**(reanneal)或**复性**(renature)形成**同源双链**(homoduplexes)。另外，单链可以与其他互补的对应物退火，如标记的核酸探针形成**杂种双链**(hybrid duplex)或**异源双链**(heteroduplex)。这种技术的能力是可以用标记的核酸探针以

高度专一性和灵敏度在复杂混合物中检测互补分子。杂交可以发生在 DNA 与 DNA、DNA 与 RNA 或 RNA 与 RNA 之间，也可以是分子内或分子外。杂交可以发生在溶液中的核酸或在溶液中与被固定分子之间(在固体支持上或固定在细胞原位上)。

杂交参数 双链核酸的稳定性取决于内在和外在因素。影响双链稳定性的内在属性反映了固定两条单链的氢键数目，包括双链长度、GC 含量和互补双链间错配的程度。双链越短、GC 含量越低、错配越多，固定两条链的氢键也就越少，它们就更容易变性。影响双链稳定性的外在属性反映了与氢键作用的环境因素。增加温度引起氢键的破坏，因此双链在温度增加时熔解(**热熔解**)。化学环境同样很重要：Na^+ 离子增加双链的稳定性，而去稳定的试剂破坏氢键。

假定双链的内在属性是不变的，所以其稳定性的维持能力被外在条件所控制，总称为**严格性**(stringency)。双链内在稳定性可以在不变的化学环境中，通过**熔解温度**(T_m)测定来度量。双链核酸的变性引起在 260nm 波长紫外吸收的变化，即可以在适合波长(OD_{260})下测定的**减色效应**(hypochromic effect)。T_m 是对应于 50%变性的温度定义的，也就是说 OD_{260}处在双链和单链 DNA 数值中间的时候。完全互补双链的 T_m 值可以通过下面的方法计算。T_m 值每有 1%的错配就下降 1 度，在甲酰胺杂交溶液中是 0.6 度。

核酸杂交实验通过测定 T_m 能用来确定两个核酸间的互补性。相反，T_m 可以用来指导杂交在精确的严格性下进行，与一些分子杂交而不与另一些分子杂交。在有些条件下，希望只检测完全互补的顺序，就用高度严格的条件。在希望检测完全互补和相关序列的其他情况下，选择低严格条件来检测特定程度的互补性。

DNA： $T_m = 81.5 + 16.6(\log_{10}[Na^+]) + 0.41(\%GC) - 500/长度$

RNA;RNA - DNA： $T_m = 79.8 + 18.5(\log_{10}[Na^+]) + 0.58(\%GC) + 11.8(\%GC)^2 - 820/长度$

寡核苷酸： $T_m = 2(AT 碱基对数目) + 4(GC 碱基对数目)$

溶液杂交 在溶液中两种核酸混合，可以研究顺序复杂性、基因组组成和基因结构。在过去，研究基因组复杂性和基因分布状态的 Cot 分析和研究转录本丰度和表达参数的 Rot 分析是溶液杂交的主要应用(见文框 12.1)，但随着基因组作图和测序计划进展这类实验废弃不用了。然而任何涉及单链核酸在溶液中退火分子反应都是利用溶液杂交的(包括对 PCR 的引物、体外突变、引物延伸、cDNA 合成和随机引发以及如消减杂交、核酸酶保护和同源多聚尾巴等技术)。

简单滤膜杂交 **滤膜或膜杂交**通过将变性的靶核酸固定化在固体支持物上改进了对杂交分子的检测。将核酸转移到支持物上的过程称为**印迹**(blotting)，这些支持物经常是硝酸纤维素膜或尼龙膜。印迹最简单形式是变性样品直接放在膜上(**斑点印迹**)。另外，靶核酸可以通过一个狭线，使被靶核酸覆盖的膜区域被确定(**狭线印迹**)。一旦转移后，核酸在膜上就不可移动。尽管尼龙膜可以自发结合核酸，但经常是通过烘烤或紫外光交联。这时将膜与含有探针的杂交液孵育，杂交进行几个小时。然后洗膜和检测探针(探针合成和检测的讨论参见*核酸探针*)。这是确定靶序列中是否存在特定序列的快速诊断技术。

Southern 和 northern 杂交 在印迹前加上了在凝胶中电泳分离 DNA 片段或 RNA 的步骤。电泳分离的 DNA 从凝胶到固体支持物的毛细管转移最先是由 Edward Southern 发明的，所以称为 **Southern 印迹**。延伸开来，类似的固定电泳分离的 RNA 的方法称为 **northern** 印迹*[1]。这些技术所依赖的原理是 DNA 片段或 RNA 分子在滤膜上的位置代表了凝胶上的位置，反映了分子量的大小。

1) 此方法是由 J.C. Alwine 等人在 Southern 印迹的基础上发展起来的，故十分风趣地称这种方法为“northern (北方)”印迹。——译者注

Southern 印迹有多种用途,可分成两组:①简单的 Southern 杂交和②基因组 Southern 杂交。简单的 Southern 用来完成克隆 DNA 的限制性图谱的研究,鉴定重叠片段和装配克隆重叠群。基因组 Southern 杂交涉及整个基因组 DNA 的消化和电泳分离。对大多数基因组而言,用 6-切割酶消化产生不同大小的上百万片段,在电泳凝胶上产生一个无法分辨的分散分布,但杂交可以确定单个片段。基因组 Southern 的一个主要应用通过限制性片段大小的改变,鉴定基因组间的结构变化(限制性片段长度多态性,RFLP)。很多人类致病突变可以通过这种方式的 Southern 印迹被鉴定。点突变可以产生或消除限制性位点,然后改变观察到的条带模式。大的缺失可以失去两个相邻的限制性位点,此后整个限制性片段缺失(参见*杂合子丢失*)。当 DNA 整合到基因组中时,可以看到限制性片段大小的增加,如通过转座因子或外源 DNA 插入到转基因小鼠中。高度变化的 DNA 中 RFLPs 分析可以分析微卫星多态性(参见 *DNA 分类*)。基因组 Southern 的第二个主要的应用是研究相关 DNA 序列的家族。一个探针不仅鉴定它同类目标,而且可以鉴定未知的其他目标,随严格性下降,被鉴定的目标数也增加。同样严格的条件可以用来筛选 DNA 文库,试图分离代表多基因家族新成员的相关克隆。同样的技术也可用于鉴定不同物种间的相关序列[参见*动物印迹*(zoo blot)]。

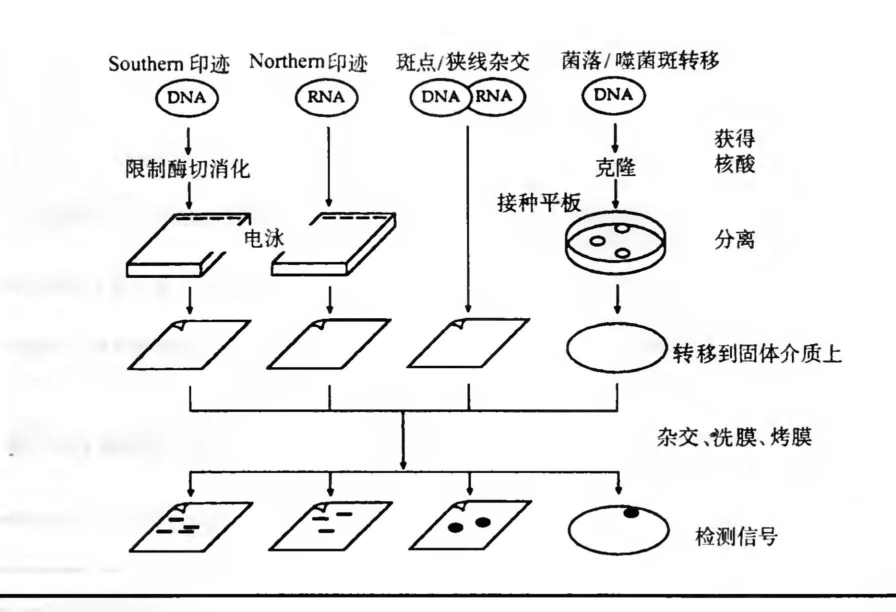
*Southern 印迹是根据发明者命名的,经常用大写字母,但 northern 印迹(RNA 的)和 western 印迹(蛋白的)是根据类似命名的,不用大写。很多人试图使 eastern *印迹*流行起来,这是近来发展的一种脂的分离和固定化技术,但这个命名未被广泛接受。还有 **southweatern** 和 **northwestern** 印迹是对 western 印迹的修改,探针是标记的核酸,用来检测核酸结合蛋白(参见 western *印迹*)。

等位基因专一性杂交　DNA 印迹(Southern 印迹和 northern 印迹)的主要应用是在分析人类疾病基因组,以**等位基因专一性杂交**。寡核苷酸探针对碱基错配极度灵敏,并且杂交条件可以被控制使单个错配导致杂交失败。这样可以用来检测由点突变产生的特殊的等位基因(**突变检测**,参见*突变筛选*)。类似的以 PCR 为基础的技术涉及不能与单碱基错配杂交的引物(参见*等位基因专一性 PCR*、*寡核苷酸连接分析和连接酶链式反应*)。

反向杂交　经典的 Southern 和 northern 杂交用一个简单的同源探针筛选包含固定靶分子的复杂混合物。**反向杂交**(reverse hybridization, **反 Southern、cDNA Southern**)涉及到相反的过程,这个过程是将克隆的 DNA 固定化与探针复合物如标记的全部 RNA 或 cDNA 混合杂交。这样的技术对快速高产量的表达研究非常有用,可以多个克隆被同时测试,如确定每个克隆的基因在特定组织中是否表达,而不需要分别杂交。

菌落印迹和噬菌斑挖取　核酸印迹的另一个例子是用来产生精确图谱的**菌落印迹**(colony blot)或**噬菌斑挖取**(plaque lifts)。在这些用来筛选 DNA 文库或分析克隆实验结果的技术中,硝酸纤维素膜或尼龙膜覆盖在平板的细菌菌落或噬菌斑上,将一些细胞或噬菌体转移到膜上。DNA 然后被变性和固定化,用特定的供体 DNA 通过杂交筛选。阳性信号对应于原始平板的菌落或噬菌斑可以移出培养,分离大量的供体 DNA。

原位杂交　**原位杂交**是核酸探针与保持在正常细胞位置的靶进行杂交。这种技术主要有三种应用——克隆基因在染色体上的细胞遗传作图(对染色体异常分析也有用;参见 FISH),检测病毒基因组和 mRNA 表达的定位。后者是确定基因表达分布的有力而快速的技术,对用抗体检测蛋白表达是一种补充。细胞或组织作为组织切片或整个标本被固定,与反义 RNA 或寡核苷酸探针渗透和孵育。非同位素探针技术的发展允许用不同的发色测定,同时分析几个基因的表达。



文框 24.4 关键工具和技术 Ⅳ:核酸探针

探针结构 **探针**是被标记的核酸,也就是以一些方式经化学修饰后与其杂交的任何物体都可以被检测到。有三种主要类型的探针:化学合成和末端标记的**寡核苷酸探针**;**DNA 探针**是克隆的 DNA 或 PCR 产物,且可以末端标记或在体外复制时进行内部标记;**cRNA 探针(互补 RNA 探针、核糖探针)**是体外转录克隆 DNA 时内部标记的探针。RNA 探针和寡核苷酸探针一般以单链分子标记。DNA 可以标记成单链,也可以是双链,但只有单链时才能作为探针发挥作用,所以必须在使用前变性。

探针标记 产生探针的不同方法在下面的表中表示。最高的**专一性活力**(每群探针中掺入标记的比例)的探针通过**内部标记**产生的,很多标记的核苷酸在 DNA 或 RNA 合成时掺入。**末端标记**涉及 DNA 3′末端加上标记核苷酸或交换 5′磷酸基团。DNA 探针也可以为特别的应用末端标记,如鉴定分子的一个末端是重要的(参见限制性图谱、DNase 足迹、转录分析)。

同位素和非同位素标记体系 通常核酸是用放射性同位素标记,如^{32}P 和^{35}S(和最近的^{33}P),可以通过放射自显影检测。这些**放射性探针**非常灵敏,但它们的操作必须有严格的安全防范,并且信号消失相对较快。最近,一系列的非同位素标记系统发展起来,它们产生发色或化学荧光信号。广泛使用的标记是**地高辛配基**(digoxigenin),是从洋地黄中分离的植物甾类。它可以与核苷酸交联并掺入 DNA、RNA 或寡核苷酸,然后用抗体检测。另一个系统是用**生物素**,它是一种维生素,细菌蛋白**亲和素**以强亲和力与其结合。生物素标记的核苷酸作为标记掺入,用亲和素检测。检测的分子可以与荧光染料或酶偶联,便于信号的检测。这些体系的优点是它们还可以用来从复杂的混合物中抽提核酸[**亲和捕捉**(affinity capture)]。生物素/亲和素体系广泛用来从复杂混合物中捕捉和抽提特异 DNA 片段(参见消减杂交、捕捉 PCR)。

挂锁探针 探针技术的最新发展是探针结构由与靶互补的两个片段组成,它们通过一个非特异的接头连接。探针的两端与靶的邻近片段杂交(末端进入片段中部),通过 DNA 连接酶连接形成

环绕靶 DNA 的拓扑结构闭合的环。这样挂锁探针(padlock probes)极其灵敏,因为锁上的探针即使在非常严格的洗脱条件下,还保持与靶序列的联结。与需要精确控制严格条件的等位基因专一性杂交(参阅)相比,阻止连接的点突变可被这种方法简单地检测到。

标记方法	探针	评论
5′末端标记	Oligo/DNA	用标记的游离核苷酸的 γ-磷酸基团由 T4 多核苷酸激酶催化取代 5′磷酸基团
3′末端标记	Oligo/DNA	用末端转移酶以标记的核苷酸加尾
缺口平移	DNA	由DNase I 在双链 DNA 中导入缺口,自由的 3′末端由 DNA 聚合酶 I 用标记的核苷酸延伸
随机引发	DNA	短的随机引物与变性 DNA 退火,由 DNA 聚合酶用标记的核苷酸延伸。比缺口平移产生更高专一性活力
引物延伸	DNA	如上,但用特异性引物。如用于热循环中标记 PCR 产物
单链合成	DNA	用标记的核苷酸由 M13/噬菌粒载体或不对称 PCR(参阅)产生单链 DNA。链专一性
体外转录	RNA	用标记的核苷酸,通过体外转录产生单链 RNA,链专一性

文框 24.5 酵母克隆载体

主要的酵母载体的分类和应用 第一个酵母载体,**酵母整合质粒**(yeast integrative plasmid, YIps),是基于大肠杆菌质粒载体的,因为很少能随机整合到酵母基因组,故转化细胞效率低。所有的酵母质粒载体都有大肠杆菌复制起始点(参见*穿梭载体*),一个或多个在酵母和大肠杆菌中起作用的选择标记(经常是筛选营养缺陷型恢复的标记)和让供体 DNA 插入的一个多个克隆位点。其他酵母质粒与 Yips 不同在于含有酵母复制起始点(除了大肠杆菌复制起始点),允许在酵母基因组外维持。**酵母附加型质粒**(yeast episomal plasmid, YEps)携带从酿酒酵母 2μ 质粒的复制起始点。这样的载体在胞质内以高拷贝维持,并高效率转化酵母细胞。**酵母复制质粒**(yeast replicating plasmid, YRps)包含 *ARS* 元件(参阅)、酵母染色体复制起始点:这些元件不稳定,在出芽时经常保留在母细胞中;偶尔它们会像 Yips 一样整合。**酵母着丝粒质粒**(yeast centromere plasmids, YCps)包含一着丝粒。它们以低拷贝存在,像孟德尔性状一样传递。**酵母人工染色体**(yeast artificial chromosomes, YAC)包含中心粒、一个 *ARS* 元件和端粒。这些大容量载体用来克隆大片段的 DNA(表 24.3),以及近来用于转基因小鼠。基于反转录转座子 Ty 的酵母载体通过转座在基因组中扩增,使整合的基因高水平表达。酵母载体的基本特性在下面的表中总结。

酵母寻靶载体 Yips 含有与内源基因同源区域,可用于基因寻靶。两种载体类型被使用。标准的 YIps 在同源区域中有单一的限制性位点,有利于单交换然后插入整合到靶基因座中。**多拷贝整合载体**是用这一原理发展起来的,通过对重复 rRNA 基因的寻靶。第二类载体(**酵母置换质粒**)促进双交换,置换了染色体 DNA 的一部分,在特定位点产生稳定的单拷贝整合。类似的原理用于哺乳动物寻靶载体(参见*基因寻靶*),在哺乳动物基因组中,尽管同源重组比随机整合事件相比要少得多。

酵母转化 DNA 可以象在细菌中一样,有多种方法被导入酵母细胞。一个有效的方法是用完整细胞在 0.1mol/L 乙酸锂悬浮,加 DNA 和聚乙二醇(PEG),接着热休克。酵母细胞也可通过电转化。

一个高效但费力的方法是去除细胞壁，产生**球芽**(spheroplasts)，在钙离子和 PEG 存在下吸收 DNA。完整细胞的方法适合于大多数应用，但球芽为转化 YAC 载体所需。

载体	组成	特性	应用
YIp/YTp	大肠杆菌质粒复制起始点	低转化效率。在细胞质中不能维持游离状态，只作为整合的元件(不能恢复)	稳定转化 具有替代遗传学的用途
YEp	2μ 质粒起复制始点	高转化效率，以大约每个细胞 100 拷贝稳定维持在细胞质中	互补功能分析
YRp	酵母染色体 *ARS* 元件	高转化效率，但因与母体细胞结合而不能稳定维持 以整合元件维持(见 Yip)	互补功能分析，或稳定整合
YCp	酵母染色体 *ARS* 元件和着丝粒	高转化效率，以低拷贝数稳定地维持在细胞质中	功能分析，特别是基因剂量效应有害时
YAC	酵母着丝粒、*ARS* 元件和端粒	如果染色体长度为 50kb 左右，则稳定维持，每个细胞 1—2 个拷贝	提供最大容量的克隆载体，具有制备大的真核基因组文库的用途
Ty	Ty 反转录转座子	在YEps 中被解离的 Ty 载体是强表达结构。Ty 载体自发整合到宿主染色体中	高效率表达

文框 24.6 植物的遗传操作

冠瘿病 根癌农杆菌(*agrobacterium tumorfaciens*)是一种革兰氏阴性土壤细菌，在很多双子叶植物中引起**冠瘿**病(crown gall disease)。细菌在土壤中感染受伤细胞(植物根冠细胞)，引起它们生长形成瘤或菌瘿。冠瘿细胞证明有两种新的属性：它们可以在缺乏植物生长激素下增生和它们合成了一个或多个修饰的氨基酸称为**冠缨碱**(opines)。

Ti 质粒 根癌农杆菌的肿瘤生成能力来源于质粒，**Ti**(tumour-induced)**质粒**，它根据不同株系其区域大小在 5～450kbp 之间变化。有两类 Ti 质粒，每种诱导不同类型的冠缨碱，**章鱼碱**(octapines)或**胭脂碱**(nopalines)的合成。一般所有的 Ti 质粒都有四个保守区域。其中最重要的是 **T-DNA**，它携带编码植物激素和冠缨碱合成的酶。冠瘿病是由植物基因组被这部分质粒以类似于细菌接合(参阅)的过程转化所致。这需要质粒的第二个区域，即含有**侵入性基因**(virulence genes)的 ***vir*** 区域。第三个区域携带与细菌间质粒的转移有关的接合基因。最后一个区域编码的功能与冠缨碱利用有关。因此通过感染和转化受伤植物细胞，杆菌将植物根冠细胞转变成生产它们能利用的特定营养的安全避难所和工厂。

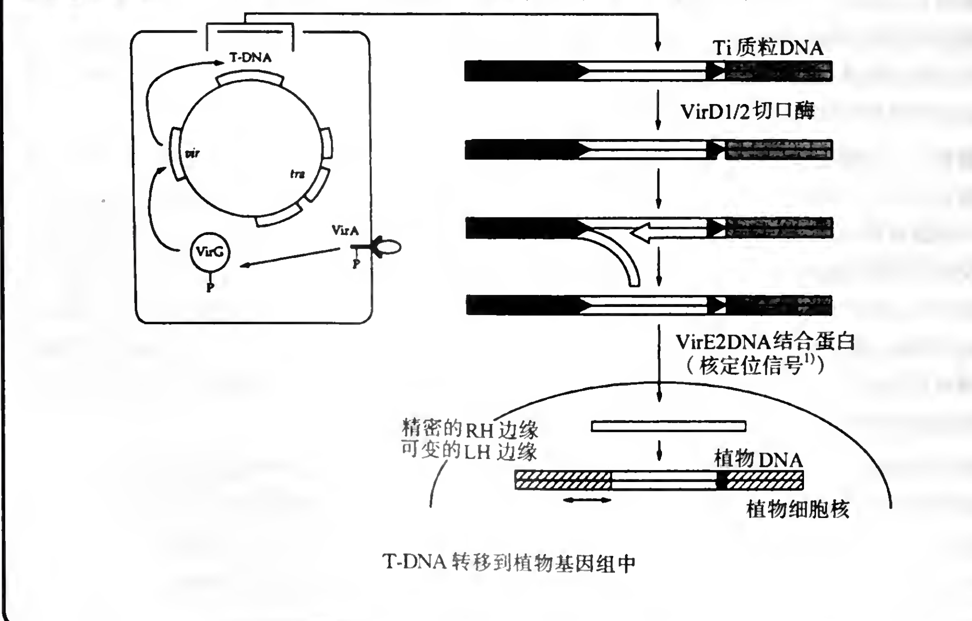
T-DNA 的结构 Ti 质粒的 T-DNA 区域是根据 25bp 不完善的正向重复序列定义的。在章鱼碱和胭脂碱质粒中的 T-DNA 结构是不同的。在前者，T-DNA 分为携带植物激素基因的和编码章鱼碱合成酶的 T_L，以及和携带合成其他冠缨碱，如农杆氨酸的基因的 T_R 两个部分。胭脂碱质粒 T-DNA 是单个片段，在其右侧有植物激素和胭脂碱合成酶基因，而在其左侧有编码其他冠缨碱合成基因。

T-DNA 转移到植物基因组 转移过程(见图)由位于质粒 *vir* 区域的侵入性基因所控制。*virA* 和 *virG* 编码对受伤植物细胞释放的酚类化合物反应调节因子。VirA 是受配基刺激自身磷酸化的

受体，并将磷酸基团转移给 VirG 蛋白，VirG 是留在 *vir* 基因座上的转录激活因子，其中每个操纵子都含多个开放阅读框架。已知几个下游 *vir* 基因的功能。VirD1 和 VirD2 形成核酸内切酶，在 T-DNA 上产生缺口并启动转移。T-DNA 以被 *virE2* 编码蛋白包被的单链形式(**T－链**)转移。VirE2 蛋白有核定位序列，负责将 T-DNA 转移到植物细胞核中。T-DNA 整合到植物基因组中有精确的右手边界，在 25bp 重复顺序的 1～2bp 位置，但可变的左侧边界可以定位在左侧 25bp 重复顺序上或在 T-DNA 内部 100bp。邻接右手边界的顺序称为**超驱动**(overdrive)，它增强转移过程并与 *virC* 操纵子编码的蛋白结合。

Ti 质粒作为植物基因转移载体　　Ti 质粒高效的转移外源 DNA 进入植物基因组的能力，使其成为植物基因转移有吸引力的载体。天然质粒不适合作为载体有两个原因：第一它们可生产肿瘤；第二它们太大不适合体外操作。肿瘤生成问题通过癌基因已被破坏或缺失的**去毒 Ti 质粒**(disarmed Ti plasmid)得到解决。T-DNA 携带的基因在转化宿主基因组的能力中不起作用，所以内部区域可以被任意操作。冠缨碱合成酶的强启动子用来表达新的转基因和筛选重组的显性选择标记。质粒太大的问题有两种方法解决。在第一种策略中，T-DNA 区域克隆在**中间载体**中操作后再导入根癌农杆菌中与内源性 Ti 质粒重组。在第二个策略中，**二元载体系统**是 T-DNA 在小的质粒中克隆和操作，而 *vir* 基因由第二个质粒反式提供。在两种情况下，T-DNA 的对植物基因组的转移都有效。

对双子叶和单子叶的转化　　转化双子叶植物的一般策略是切割叶片(引起细胞损伤)，与携带重组消除 Ti 载体的根癌农杆菌孵育。叶片转移到茎干－诱导培养基中，同时对 T－DNA 携带标记进行筛选。几周后，茎干生成后转移到根－诱导培养基中，再过几周后可以转移到土壤中。尽管 Ti 质粒为基础的转化系统是基因转移的有效和广泛方法，但它限制在对感染敏感的双子叶植物中。为了对其他植物进行操作(包括主要的谷类：大米、小麦、玉米)，发展了另一些技术。原生质体的电穿孔适用于这些物种，尽管这些物种并不都能从单个细胞中重新生成。应用最广泛的技术是显微弹转染，可以用来转化双子叶和单子叶植物，包括大米、小麦和玉米。



T-DNA 转移到植物基因组中

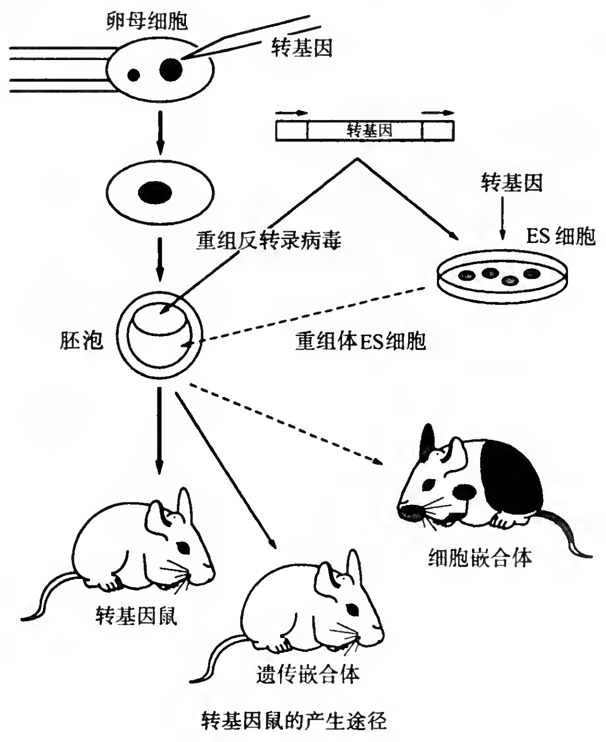
1）核定位信号(nuclear localization signal)，原文误为 nuclear localisation signal。——译者注

文框 24.7　转基因动物

动物胚系细胞转化路线　与植物细胞不同，分化的动物细胞不能重新产生整个有机体。转基因动物必须从全能性细胞的操作产生，如卵细胞和非常早期胚胎的细胞。一般，三种形式的遗传操作被用来产生转基因动物（见图）：①DNA 微注射；②重组反转录病毒感染；③对胚胎干细胞的转化。所有三种途径都可以用于产生转基因啮齿动物，但微注射是其他转基因动物的主要途径。

DNA 显微注射　为了通过这种方法得到转基因小鼠，DNA 直接显微注射到刚受精的卵的雄前核中（选择雄性前核是因为它比雌性前核大）。受精卵是从解剖交配不久的超排卵雌性动物的输卵管中取出。只需将转基因注射（即不需要载体）。存活的胚胎被培养直到胚泡期被植入假孕的代孕雌性动物中。DNA 的整合经常以串联排列发生（**转基因组**），通过非法重组，注入 DNA 片段的产生末端连接。经常只有一个整合位点，转基因组以孟德尔方式传递，大到 50% 的后代可以是被转基因的。偶尔，直到一个或两个早期细胞分裂后，整合才发生，在这种情况下小鼠是遗传嵌合体。随机整合位置和拷贝数意味着转基因经常服从于位置和基因剂量效应。它们偶尔也可能破坏内源性基因。

卵细胞或早期胚胎的显微注射用来产生转基因牛、山羊、绵羊、猪、非洲爪蟾（*Xenopus*）、鱼（如斑马鱼，*zebrafish*）、蠕虫（*Caenorhabditis*）和果蝇（*Drosophila*）。在果蝇中，胚系转化是通过重组体缺陷 P-因子介导的，P-因子被克隆在质粒中，与第二个含有顺式缺陷型（剪翅）P 因子（参阅）的质粒一起注入胚胎，也就是说一个在反式提供转座酶但不能移动其本身。如果将重组的 P 因子导入 M-细胞型的胚系中（其缺乏内源性 P 因子，因此没有内源性转座酶活性），转基因稳定整合到基因组中（参见杂种不育）。



转基因鼠的产生途径

反转录病毒转导　植入前的小鼠胚胎暴露于反转录病毒中，经常将一个拷贝的原病毒整合到一个或多个细胞中。携带转基因的重组缺陷病毒这样可以将外源 DNA 整合到小鼠的基因组中。与原核显微注射方法相比，反转录病毒有几个缺点。转基因 DNA 必须克隆在容量限制在 8～9kbp 的合适的载体中，整合需要辅助病毒（除了预防措施，它可能导致在转导后病毒增生，引起任何基因的破坏），因为整合原病毒的重新甲基化，转基因可能不表达（见 DNA 甲基化及外遗传调节）。反转录病毒因此对有些应用有效，但对制备转基因小鼠不是万能的。

胚胎干(ES)细胞　第三个产生转基因小鼠的方法涉及全能性**胚胎干细胞**(embryonic stem cells)的使用，胚胎干细胞是从早期胚胎内细胞团衍生的无限增殖化细胞。这样的细胞可以用裸露 DNA 转染或重组反转录病毒转导，并且转化的细胞可以注入囊胚泡，与胚胎合并参与到发育的小鼠中。ES 细胞方法优越性有几个理由：转移到胚胎前，可以选择成功转化的细胞，所以成功率较高，并且 ES 细胞可以用来检测很少的同源重组，适合产生定向突变小鼠（见正文）。ES 细胞注入囊胚泡产生细胞嵌合的小鼠，所以用来作为生育的母小鼠必须在胚系转基因转移前交配。ES 细胞优点还在于性别和毛皮颜色可以用作细胞和遗传标记。如果从小鼠 129 株（有显性野鼠色毛皮表型）取得的雄性 ES 细胞注入雌性 C57B10/J 母鼠（隐性黑表型）的囊胚泡中，那么第一代小鼠是细胞嵌合的，也就是说它们根据每种细胞系的克隆分布有不同颜色的斑纹。转基因的胚系转移可以通过第二代的野鼠色表型而被建立，成功的转移通过嵌合雄鼠与黑色母鼠的交配系增加，因为雄性的 ES 细胞可以有产生嵌合的生殖组织。

文框 24.8　*基因治疗*

基因治疗的范围　重组 DNA 技术以很多针对人类疾病新方法丰富了医学领域。除了提供新的诊断方法和动物疾病模型以外，从克隆和操作人类基因直接提出了很多新的治疗策略。这包括表达克隆人类基因产物，发展新的疫苗和治疗性抗体的工程化。一种涉及用 DNA 治疗的完全不同的方法可用来减弱疾病。这样的**基因治疗**(gene therapy)可涉及患者体内细胞的遗传修饰（**体内基因治疗**）或对培养细胞的遗传修饰，在重输给患者（**体外基因治疗**）。体内基因治疗同时包括了目标细胞的遗传修饰（体细胞转基因）和用 DNA 作为基因组外处理的治疗（不改变核苷酸顺序）。这样的治疗能用来缓解或纠正人类基因突变引起的症状，或阻止感染性疾病（如干扰病毒的生命周期）。

基因增大治疗　常规的基因治疗方法是将 DNA 加到基因组中，以替代缺失的功能，这被称为**基因增加治疗**(gene augmentation therapy, GAT)。基因增加治疗的目的是通过加入一个有功能的等位基因，纠正功能失去的效应，如有缺失引起的。转移的基因可以稳定地整合到基因组中（在这种情况下，可能可以永远校正缺陷，特别是干细胞被转化）或者以游离在细胞质中的形式维持（在这种情况下，不可避免有基因表达的下降，治疗需要反复进行）。几种 GAT 临床试验正在进行，包括囊性纤维化、腺苷脱氨酶缺陷和家族性高胆固醇。

在核酸水平的基因抑制治疗　为了治疗显性失活功能缺失突变和功能获得突变，基因增加治疗就不是那么有效：加上功能失常基因的功能性拷贝对表型不会有影响。按理治疗这样疾病的有效方法是**定向纠正**（等位基因替代）或基因剔除来去除突变的等位基因。然而在实际上这是效率低下的过程，即使在培养的细胞中，将其应用到在很多体细胞，特别是在体内，纠正遗传缺陷还有待

于技术的进一步改进。一个有望将来在基因治疗中发挥作用的新方法是用核酶或 RNA 编辑酶在 RNA 水平纠正致病性 mRNA。

在另一个目前正在对多种类型癌症临床试验的策略中,是用核酸抑制基因表达。这种方法的优点是抑制物可以定位抑制特定等位基因,所以正常功能的等位基因表达不受影响。反义基因的导入能产生稳定和不断的反义 RNA 的表达,与(突变)mRNA 结合并抑制其翻译(它也可以定向地将突变 mRNA 降解)。更进一步的是用含有核酶的反义结构,使所结合的 mRNA 降解。寡核苷酸可以被用作基因组外的基因治疗(如不涉及改变基因组的治疗)并且可以以两种方式进行。反义寡核苷酸与反义 RNA 一样起作用,通过与 mRNA 结合引起抑制和降解(在这种情况下,可以通过消化 DNA:RNA 杂种中 RNA 链的 RNaseH)。第二,嘧啶丰富的寡核苷酸能够与嘌呤丰富的 DNA 链形成三链结构(通过 Hoogsteen 碱基配对,参阅),抑制转录(**三链治疗**)。**肽核酸**(peptide nucleic acid,**PNA**)是 DNA 类似物,它的磷酸骨架被中性的肽链骨架所替代。PNA 寡核苷酸可用作探针和基因治疗的试剂。它们形成非常稳定的三链结构。

在蛋白水平上的基因抑制治疗　定向抑制基因表达可以在蛋白水平上发挥作用,通过在细胞内表达遗传工程的抗体(**内在抗体**),抗体结合并使突变的蛋白失活。抗体并不是发展用于蛋白水平基因治疗的唯一分子。任何以多聚体发挥作用的蛋白都是显性失活抑制蛋白的潜在目标,这种策略可以被用于阻止病毒衣壳蛋白组装,包括 HIV 的蛋白。一个新的进展是用简并寡核苷酸鉴定与蛋白相互作用的特定序列。这些寡核苷酸或 **aptamers**,可以用来灭活特定的突变蛋白。

基因转移和表达的方法　最有效的基因转移载体是哺乳动物病毒为基础的,特别是反转录病毒、腺病毒和腺伴随病毒(adeno-associated viruses)(整合到基因组中,可以介导遗传缺陷的永久纠正)和疱疹病毒(神经营养性和游离型的)。最直接的基因转移方法包括注射(如肌肉)和成功产生转基因动物和植物的显微注射技术。一个包括脂质体包装 DNA 在内的转移过程在癌症基因治疗中得到广泛应用,尽管效率较低。

治疗基因表达的专一性可能是疾病控制或纠正的重要因素。转移的基因可以与细胞类型专一性的启动子相连,但这种调节由于受整合基因内源性影响较弱。专一性可以通过特定细胞以某种方式定向达到。直接寻靶对特定类型细胞是可能的,通过利用病毒的营养特性,DNA 与细胞类型专一性受体的配基接合(通过受体介导内吞吸收 DNA)或通过控制转移位置(注射和显微注射方法)。汽雾也被用于将重组病毒载体导入肺中(如对囊性纤维化的治疗)。

(陈　淳译)

第25章
重　　组

基本概念和定义

- **重组**(recombination)是已经存在的遗传物质产生新的组合的过程。**分子间或染色体间重组**通过混合离散染色体产生新的组合(真核染色体减数分裂时独立分配和病毒基因组片段的重新分配),而**分子内或染色体内**重组是酶依赖的过程,新的遗传物质通过DNA的剪切和连接产生。有五种类型的分子内重组(表25.1),尽管每种情况中分子反应——DNA剪切、双链间的链交换、DNA修复和去除——是相似的。
- 重组和突变被认为是遗传变异的两种截然不同的机制:重组是已经存在的信息重排,而突变是在基因组中导入新的信息。尽管这些定义十分有用,突变和重组在分子水平上是相互关联的。很多重组事件(特别是引起转座和不正常重组)会导致基因破坏,被描述成突变。相反的,一些重组为修复潜在突变或致死DNA损伤所必需。
- 重组以两种方式被应用:对遗传座位作图和基因及基因组的操作。遗传作图根据染色体上两个基因座距离越远越容易发生同源重组的原理进行。因而一个特定杂交的重组产物的比例提供了对物理距离的估测(见基因结构和作图一章)。同源重组可被用于基因寻靶,而位点专一性重组可以诱导缺失和染色体重排。转座和不正常重组也被用来基因转移和整合(见重组DNA和分子克隆,可移动的遗传因子各章)。

表25.1　分子内重组的不同类型

同源重组

重组对之间需要有同源性。调节这一过程的蛋白(如大肠杆菌中的RecA)不是序列专一性的,而是同源依赖的。经常涉及长的同源区域(如减数分裂中)。

位点特异性重组

重组对之间不需要同源性。调节这一过程的蛋白(**位点特异性重组酶**)在供体和受体分子中识别短的,特异DNA序列,这些蛋白之间的相互作用帮助重组。供体和受体位点之间经常存在同源性,因为同样的重组酶蛋白可结合两者的识别位点。

转座

重组对之间不需要同源性。调节这一过程的蛋白(**转座酶,整合酶**)识别重组分子中的短特异序列,这样的序列称为转座因子(识别位点经常是转座因子和宿主DNA之间的接合处)。受体位点一般在序列上相对非特异,重组将可转座因子整合到宿主DNA中(见第13章可移动的遗传因子)。

不正常重组

重组对之间不需要同源性或少量同源性,是异常细胞作用的结果。包括复制中不正常末端连接,链滑动或成环。**不等交换**(参阅)经常被描述成不正常,尽管其机制是正常的(但在错误的地方发生)同源重组。

人工重组

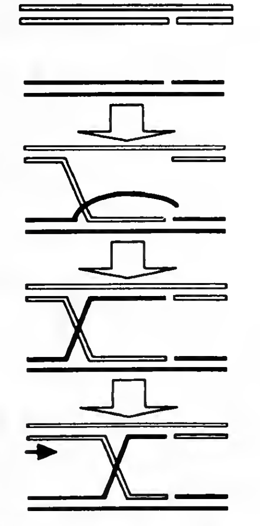
体外用纯化的酶和底物进行的**DNA连接**(参阅)引起的重组(见重组DNA一章)。

25.1 同源重组

同源重组——作用和机制　同源重组是同源性依赖的，但不是序列依赖的，所以任何两个具有相关序列的 DNA 分子可以通过这个过程进行重组。同源重组有两个基本功能，遗传混合和 DNA 修复，尽管这一过程首先行使哪一种功能还不清楚。其他加工也依赖于同源重组：转座因子的被动运动(见第 13 章可移动的遗传因子)、特定噬菌体的复制(见病毒一章)、酵母交配类型的转换和相关表型(见下文)，以及减数分裂时染色体的分配(见下文)。同源重组被用于真核基因组的遗传作图(参阅)(见基因结构和作图)和基因寻靶(参阅)——将 DNA 人工导入细胞，将一个等位基因用另一个替代(见第 24 章重组 DNA 与分子克隆)。

同源重组的模型分成三类：**拷贝-选择模型**，重组在 DNA 复制中发生：新生链延伸时转换为新的模板。**断裂和重新连接模型**中，重组在没有 DNA 复制时发生：DNA 链打断在双链间交换和重新连接。还有杂合模型是结合了上面两种模型的特征。所有三种类型的重组在不同的环境中发生。减数分裂时发生的同源重组基本上是通过链断裂和重新连接，而通过重组的 DNA 修复涉及拷贝选择和杂合机制。

同源重组的步骤　同源重组分成四个阶段：联会、链转移、修复和分解。联会和链转移在图 25.1 中显示。



(a)：联会。同源双链配对。

(b)：起始。缺刻链从起始双链侵入到受体双链中，置换原有链。

(c)：链的交换。被置换链填补到起始链的缺口中交换点即霍利迪连接被建立。

(d)：分支移动。通过双链间链的交换霍利迪连接可从它原来的位置移开，根据移动的方向可增加或减少异源双链的量(图中显示分支以箭头所示的减少异源双链量的方向移动)。

图 25.1　完整双链间同源重组的**霍利迪模型**(Holliday model)，显示了联会和链转移阶段。经典的模型包括对等位置有缺口的双链。然而这不是经常发生的情况——正常情况下，只有双链中一条链是有缺口的，侵入的链以拓扑封闭单位取代对应的链——形成 **D 环(取代环)**。重组也可以发生在两条完整的双链中，这种情况需要拓扑异构酶使交换的链与它们互补的分子缠绕(参见平行接合、相缠接合)。

(1)在**联会**中，同源双链排列。

(2)在**链转移**中，一条 DNA 链从一条双链转移到另一双链中。当第一条链进入同源双链(如果受体双链是完整的)并取代停留的链时，标志着重组的**起始**。这样的过程可以产生一个短区域的**异源双链 DNA**：含有从不同亲本分子来源链的双链 DNA，可能包含有

不配对碱基,反映了亲本链的序列差别(不同的等位基因)。如果受体双链是完整的,取代的停留链能够与起始双链的自由链配对。两条转移链相互交换,形成称为**交换桥**、**交换分支**或**霍利迪连接**的结构。霍利迪连接的位置可以通过不断的双链间交换,相对起点位置发生移动。这就是**分支移动**,可以增加或减少异源双链 DNA 的数量。

(3)**修复**和**解离**在图 25.2 和图 25.3 中显示;它们不以固定的顺序进行,因为这依赖于重组对伙伴和能否得到合适的酶。修复指三个不同的过程。在最简单的例子中,重组双链是完整的(也就是说任何双链都没有遗传信息的丢失),修复涉及打断链的重接。这就是**保守重组**。霍利迪连接可以以两种平面被解离,如图 25.2 所示,产生两个产物中的一个——非重组背景中的一**小段**(patch)异源双链 DNA,或者有重组标记的异源双链的接合。

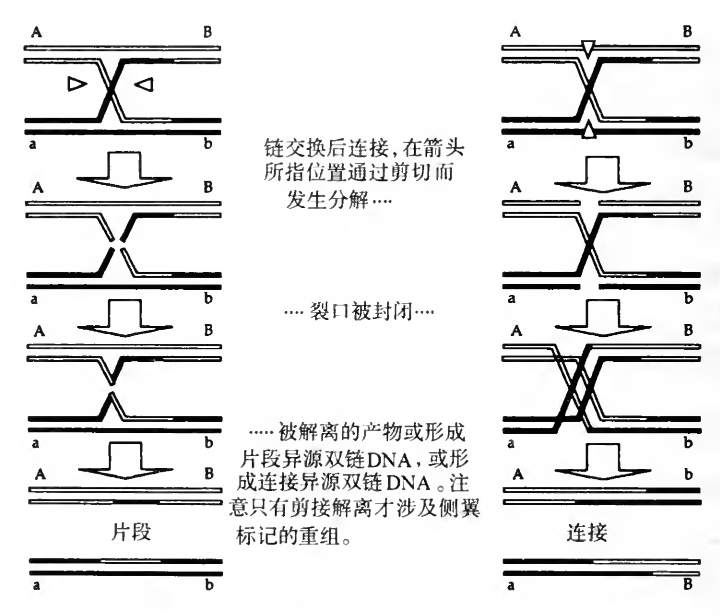


图 25.2 在两个平面上霍利迪连接的分解,产生不同的产物。只有一种分解途径产生有标记 A 和 B 的重组分子,尽管两种途径都产生异源双链 DNA,称为**联片**(splatch)[是片(patch)和剪接(splice)的组合]。

然而,如果双链中的任一条丢失了遗传信息(即如果有单链缺口或断裂),DNA 修复合成用来自同源双链的信息为模板来取代丢失的信息(图 25.3)。含有新 DNA 合成的重组称为**不保守重组**。在极端的例子中,整个染色体片段丢失,霍利迪连接的解离产生了一个能复制丢失片段的复制叉(见第 14 章诱变和 DNA 修复)。第三种类型的 DNA 修复为异源双链 DNA 错配修复,它的方向通常是随机的(参见复制后错配修复),可能引起基因转变(见下文)。

同源重组的分子基础　大肠杆菌中负责重组所有主要阶段-联会、链交换、分枝迁移和霍利迪连接、解离的蛋白已经被发现(文框 25.1)。对于重组来说还需要单链 DNA,它看来由多个重叠的途径来提供。

RecBCD 途径是大肠杆菌重组底物的主要来源。基因 *recB*、*recC* 和 *recD* 的产物结合产生大的复合物,**RecBCD 酶**或**外切酶 V**,它有三种活性:①ATP 依赖的外切酶活性;②

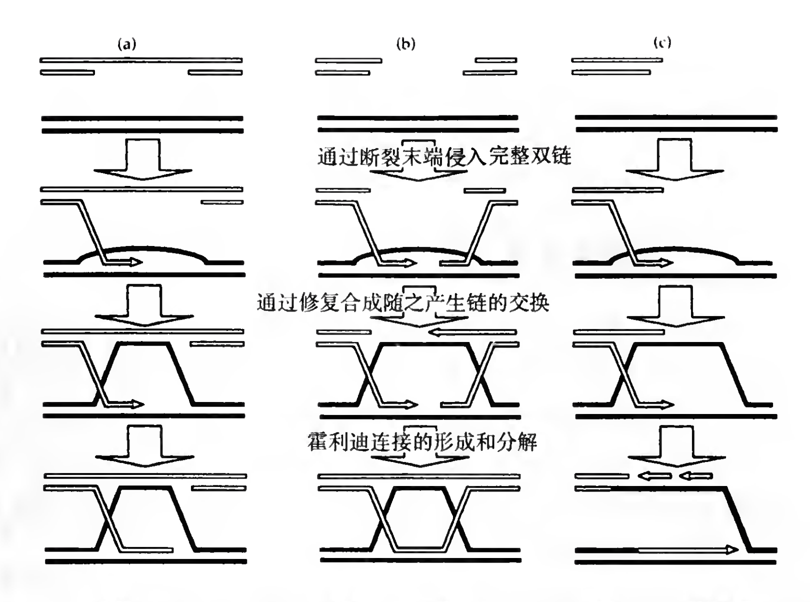


图 25.3 作为修复损伤 DNA 机制的同源重组。单链缺口 (a)和双链断口(b) 的修复涉及用未损伤双链的一条链为模板新复制的链交换损伤链。复制完成后接着进行链连接、形成如图 25.2 所显示的霍利迪连接。重组包括一个(c)产生霍利迪连接中间体的部分染色体,这个中间体形成复制叉。箭头表示新 DNA 合成的方向。

ATP 增强的内切酶活性;③ATP 依赖的解旋酶活性。

称为 **chi**(GCTGGTGG)的顺式作用位点促进重组,它们在大肠杆菌基因组大约每 5kbp 出现一次,代表了内切酶倾向性的剪切位点。RecBCD 以方向依赖的方式在 3′端大约 5 个碱基位置剪切含有 chi 序列的链。chi 位点因此是重组热点(参阅)。平末端的线性双链 DNA 是解旋酶活性的底物。酶与 DNA 结合,不断解链,在它经过的地方产生单链环。这些环偶尔被酶的内切酶活性剪切,接着进一步解旋产生重组所需要的单链尾巴。

尽管 RecBCD 途径对重组是重要的,*recBC* 突变保留高达 10%的同源重组活性,表明存在着其他途径产生重组 DNA。另外,通过在其他基因座的突变,如 *sbcA*、*sbcB* 和 *sbcC*,几乎可以完全恢复 *recBC* 突变的重组能力。分析带有基因型 *recBC*, *sbcBC* 的突变株有助于发现进一步的重组基因。它们有重叠的功能,总称为 **RecF 途径**。通过研究 *recBC sbcA* 突变体发现另一个系统,涉及 *recE* 基因的上升调节。这些重组基因和它们的可能的功能在表25.2中列出。其中很多是SOS诱导的(参见SOS反应),说明它们在细胞中的

表 25.2 大肠杆菌中涉及重组 DNA 产生的重组基因

基因	功能
RecBCD 途径	
recB、*recC*、*recD*	RecBCD 是 ATP 依赖的外切酶和解旋酶(外切酶 V),它与双链断口结合,解开 DNA 并导入裂口,其倾向性位点称为 chi(参见重组热点)。RecBCD 酶产生 RecA 可以作用的底物
RecF 途径	
recF、*recO*、*recR*	帮助链配对
recN	在双链断口修复中起作用
recJ	5′→3′单链核酸酶
recP	不清楚
recQ	解旋酶
RecE 途径	
recE	外切酶Ⅷ
recT	与 RecE 协同作用的同源链配对蛋白

正常功能是重组介导 DNA 修复。有些用于分子克隆的细菌株(参见第 24 章重组 DNA 与分子克隆)所有重组基因是缺陷的,这是为避免在大质粒中或在质粒和染色体间发生重组。

25.2 同源重组和遗传作图

连锁 **连锁**(linkage)这个术语在遗传学上有三种不同含义。首先表示卷曲的双链 DNA,用来确定双链 DNA 拓扑学参数(见核酸结构)。第二连锁表明当一个特定基因(延伸一下,联系到它们相关的表型)在特定染色体上被发现,如 X 染色体上基因和它们控制的性状称为是 X-连锁的(参见性连锁遗传)。最后,**遗传连锁**指倾向于作为一个单位遗传而不是独立分配的。这反映了紧密的物理连锁,也就是说这些基因彼此物理连接处于同一染色体上。与遗传连锁相关的名称在表 25.3 中定义。

表 25.3 与遗传连锁相关的一些术语

术语	定义
连锁	两个基因座位于同一染色体上的紧密物理关系,可阻止不相关的分配。
相位	双杂合等位基因之间的连锁特点。如果两个显性等位基因和两个隐性等位基因一起分离,这两个基因座的相位是**相引**(coupling)(或**顺式构象**),或者如果一个基因的显性等位基因与另一个的隐性等位基因一起分离,称为**相斥**(或**反式构象**)。
单倍体型	一条染色体上等位基因的特定顺序和构象。
同线性	(1)**同线性基因**位于同一染色体上。它们可能或不可能被证实是连锁的-较广泛分散的基因受到多次交换(参阅)的影响,给人们的印象是不连锁的。 (2)**同线性**是描述在不同物种中染色体上基因以相同顺序排列的术语。不同物种对应的染色体片段是**同线区域**(参见比较基因组)。
连锁群	证明是连锁并在一起作图的一群基因。在物理名称上,一个连锁群等价于一条染色体。
连锁平衡/不平衡	在一个群体中两个或更多的多态性基因座,配子中存在所有等位基因的组合称为**连锁平衡**(linkage equilibrium)。在单个基因座上的一对等位基因在随机交配群体中一代后达到平衡。在这个群体中当同时考虑两个或更多个基因座时,如果等位基因的特定组合比其他组合更普遍这就开始存在**连锁不平衡**。如果基因是不连锁的,平衡可以通过多代建立,如果基因是连锁的则需要更长的时间,因为这依赖于它们之间的交换。因此,两个基因座越近,不平衡的时间也越长。在自然种群中的连锁不平衡表明最近产生了突变或对特定单倍体型的选择。
假连锁	在不同染色体上的基因座表现类似连锁的情况。这在易位杂合子中可以发现,当一四倍体物种进行二倍化时,四个相关染色体形成不同的同源配对中也可以观察到。

真核中的连锁和减数分裂 在有性生殖中,二倍体细胞为了使染色体数目减半和产生受精所需的单倍体生殖细胞,进行特殊的分裂形式——**减数分裂**。减数分裂涉及两轮分裂中间没有插入 DNA 复制周期(参见减数分裂),第一次分裂是**还原分裂**(reduction division),这时同源染色体分离到纺锤体相反两极。第二次分裂与有丝分裂相同,姐妹染色单体分离。第一次减数分裂与有丝分裂有几方面的不同。

(1)减数分裂Ⅰ经常具有延长的前期,根据染色体的行为和外观分成五个阶段——**细线期**(leptonema)、**偶线期**(zygonema)、**粗线期**(pachynema)、**双线期**(diplonema)**和终变期**

(diakinesis)[1)]。染色体在细线期第一次变得可见。

(2)在偶线期阶段,同源染色体成对排列。这个过程称为**联会**(synapsis),对它了解得还很不清楚,它涉及一个在两对姐妹染色单体间形成的蛋白质**联会复合物**。联会可能是在端粒(参阅)起始的含有两个同源染色体的四链结构称为**二价体**。

(3)在大多数生物中,同源重组发生在联会的染色体之间,这是正确分离所必需的。这个过程发生在粗线期阶段,涉及含有重组酶的被称为**重组结**(recombination nodules)的大蛋白复合物。

(4)减数分裂Ⅰ的分离机制是独特的,涉及联会[2)]的解离而不是着丝粒的分裂。联会复合物在双线期被破坏,同源染色体因为**交叉**(chiasmata)(交换发生的位点)保持联系。在终变期**去联会**(desynapsis)持续进行,染色体变得完全浓缩。染色体排列在中期板上,重组中间体的解离是进入后期的标志,在此时同源染色体已分离。

如果两个标记位于分离的染色体上(也就是如果它们不是物理连锁的),根据孟德尔第二定律(参阅)它们独立分配。这可以在双杂合体中观察到,这种情况下等位基因的**亲组合**和**重组组合**的频率是相同的(图 25.4)。相反,如果两个标记位于同一染色体上,在每条同源染色体等位基因会作为一个单位分离,因为它们物理连接在一起。可预期在所有的减数分裂产物中都发现亲组合的等位基因,但这样的**完全连锁**(total linkage)很少发生,这是因为在减数分裂前期联会的同源染色体间经常发生交换。**交换** crossover 就是染色体断裂和链互换——分子术语中的霍利迪连接——导致亲本染色体之间的重组,从而产生等位基因的同源重组(图 25.4)。

连锁和遗传作图　减数分裂中同源重组的发生绝大部分是随机的——这种表述过于简单化,但一般是适用的(参见重组热点和冷点)——因此任何两个杂合标记间发生重组的可能性越大,它们在染色体上的距离越远。重组频率(参阅)以特定杂交中重组后代的比例测定,因此反映了标记间的物理距离。利用减数分裂对真核基因座作图的原理在其他部分详细讨论(见基因结构和作图),尽管由于克隆技术日益进步,遗传图谱的使用逐渐被大力进行的物理作图方法所替代。连锁可以用于以其他方式对基因作图,如参见随机杂交作图,HAPPY 作图,中断杂交作图,共转化和共转导作图。

有丝分裂重组　尽管真核生物的同源重组经常发生在减数分裂(当同源染色体配对时),联会和交换也在其他时期发生,特别是在如果蝇这样的物种中,同源染色体配对与体细胞相关。任何不在减数分裂中发生的重组(也就是说在没有有性生殖的情况)称为**准性交换**(parasexual exchange),这主要指是在细菌中遗传交换的形式(见第 10 章细菌中的基因转移)。在真核生物中的准性交换发生在细胞器基因(参阅)中和体细胞的细胞核染色体间(**有丝分裂重组**)。

同源染色体间的有丝分裂重组可以导致**有丝分裂分离**,杂合基因座分配到子代体细胞中,产生根据它们不同表型确定的受体同源克隆。第一个有丝分裂重组的证据来自果蝇,受体同源组织的邻近片段[**孪生斑**(twin spots)]在杂合背景中观察到。如同其他异常

---

1)前期减数分裂可分为五个阶段,细线期,偶线期,粗线期,双线期和终变期。细线,偶线,粗线,双线这些术语是广泛使用的,而上述的这些术语有严格的限制。

2)此处应涉及联会染色体的解离,原文为涉及重组解离(resolution of recombination)可能有误。——译者注

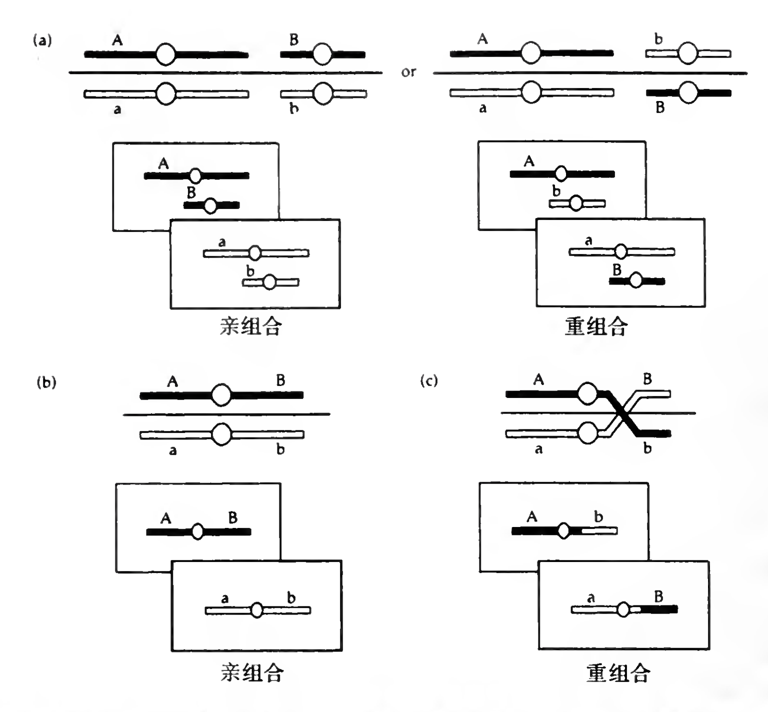


图 25.4 连锁和染色体分离。同源染色体在减数分裂前期配对(父本染色体以黑色表示,母本染色体以白色表示)。每个染色体上只有一条链在这个阶段显示出来,尽管这时有两个姐妹染色体,每个携带一个双链 DNA 分子。进行减数分裂的个体在 A 和 B 两个基因座上是杂合子,亲本单倍体型是 AB(父本)和 ab(母本)。方框代表子代细胞。(a)如果基因位于两条染色体上,它们会独立分配。亲组合(AB,ab)和重组合(Ab,aB)的等位基因组合以相同频率出现,反映了中期板(细线)上染色体配对的随机性。(b)如果两个基因位于同一染色体上,等位基因的亲组合预期是100%,因为每个亲本单倍型是物理连接的(连锁)。然而,因为联会染色体的同源重组(c),可能产生等位基因的重组合。

的染色体行为,有丝分裂重组可由 X 射线的照射产生双链 DNA 断裂诱发。说明这个过程可能是 DNA 修复的副产品。同源染色体间的有丝分裂重组能够用来进行遗传作图。这与减数图谱相类似,但对遗传距离的计算不完全一样,表明了至少推动减数和有丝分裂重组的因子是不同的。这被在酵母中发现的减数分离或有丝分裂重组基因所证实(文框 25.1)。

第二种有丝分裂重组的形式不是发生在同源染色体之间,而是姐妹染色体之间——**姐妹染色单体交换(SCE)**。这也是为 DNA 损伤所促进的,且它相对上面讨论的非姐妹染色体交换占优势,反映了细胞周期后期相关姐妹染色体更近的物理关系(参见**花斑染色**、**DNA 修复紊乱**)。

25.3 随机和程序化非交互重组

非交互重组　**交互重组**(reciprocal recombination)是对称和保守的,也就是说信息在双链间交换,没有丢失或获得。因此如果重组在携带等位基因 AB 和 ab 的亲本染色体间发生,交互重组是 Ab 和 aB。**非交互重组**(nonreciprocal recombination)涉及信息的单方向传递。非交互重组的例子是上述的杂交将产生 Ab 和 AB,信息单方向从 A 传递到 a 等位

基因，*a* 等位基因丢失了。这种类型的重组也称为**基因转变**(gene conversion)，因为它看上去是一个等位基因转变成另一个。非交互重组或基因转变在两种环境中发生。

第一，考虑在单链侵入和分枝迁移中产生的异源双链 DNA 的命运，当霍利迪结构解离，含有错配碱基的异源双链 DNA 可以成为错配修复(参阅)的目标。修复可以是两种方向的，因为减数分裂链交换发生复制后，细胞不再分辨亲本链和子代链(参见复制后错配修复)。要发生基因转换事件，基因座必须位于分枝迁移区域内的链交换位点附近，这个现象经常在重组紧密连锁的标记间观察到。因为修复是无方向性的，基因转变可能因为处于各个修复过程中具有相同机率的大量减数分裂产物中无法辨识。而在真菌代表单个减数分裂中的四分染色体中很容易观察到(参见四分体分析)。

第二，基因转变发生在缺口和断裂的重组修复中。单链缺口可能在修复前就被同源双链中一条链补平(图 25.3)，这样，如同上面讨论的，产生在两个可能方向上被修复的异源双链。双链断裂经常是外切酶的靶位点，因此在修复前可能有信息的丢失(图 25.3)。在这种情况下，基因转变经常是同方向的，也就是说向同源(未损伤的)染色体方向，因为这是唯一可获得的信息。在 DNA 被环境中的试剂损伤的情况下，染色体断裂可被基因转变来修复，转变也是遗传重排的一种机制。可保守移动的转座利用基因转变增加它们的拷贝数目(参见被动转座、移动内含子)。基因盒的开关，如特殊的酵母交配型或免疫球蛋白结构也是以这种过程形成的(见下文)。

信息盒转换　　非交互重组可以程序化发生来改变在特定基因座出现的信息。这种基因调节的机制称为**转换**，它需要一个信息能表达的活性基因座和多个具有不同形式信息的称为**盒子**(cassettes)的沉默基因座，在沉默基因座中，信息以结构性不表达的状态储存。沉默基因座和活性基因座之间的非交互重组可以由活性基因座中起始重组修复和基因转变的双链断裂所诱发。

盒子转换是细胞指导的被动转座的一种形式。它不同于活性转座和其他形式的信息交换，如细菌整合子中用不同的抗生素抗性盒子的替换，此是由于重组机制的不同。在活性转座和整合调节中(涉及位点专一性重组)，盒子整合的基因座起始是空的。在转换的情况中，信息总是存在于活性基因座中，也就是说转换是一个取代过程而不是整合过程。

酵母的交配型转换　　酵母中的交配型转换是一种简单的分化形式(参阅)。在酿酒酵母中，两种交配型称为 a 和 α。交配类型的表型反映了细胞表达的特定的信息素和受体，由 *MAT* 基因座决定，*MATa* 或 *MATα* 等位基因以相互排斥方式表达。在同宗配合的酵母中(它们高频率发生交配型转换)，一个位于 *HO* 基因座的功能性等位基因编码具有长的靶位点的内切酶，在 *MAT* 基因座特异性剪切 DNA，产生一个双链断裂。*HO* 基因受到 SWI5 调节，SWI5 是以细胞周期依赖方式的合成的转录因子，这就把交配型转换与细胞周期调节联系在一起(见细胞周期)。重组可以在 *MAT* 基因座以及位于 *MAT* 左侧与右侧几千 kb 的沉默信息盒子 *HMLα* 和 *HMRa* 中之间发生。*MAT*、*HMLα* 和 *HMRa* 的结构是相似的，在图 25.5 中显示。有四个同源区域，a 和 α 等位基因只在含有调节和编码信息的 Y 区域有差别。24bp 的 HO 内切酶靶位点处于 Y 和 Z 的交界处，在沉默交配型盒子和 *MAT* 基因座中是相同的。沉默的机制并非沉默交配型盒子的内在性质，而与位于沉默区域两侧可调节染色质结构的 **SIR 蛋白**的靶位点有关。

转换由 HO 诱导的双链断裂起始，非交互重组涉及 *MAT* 和沉默盒子之一。异源重组有 90% 的发生率，表明 Y 区域所采取的结构表现出底物倾向性。破坏沉默的突变使沉默盒子在基因转变中起到受体的作用。因此基因沉默和沉默交配型盒子的异源染色体化的机制也负责保护它们免受内切酶作用。

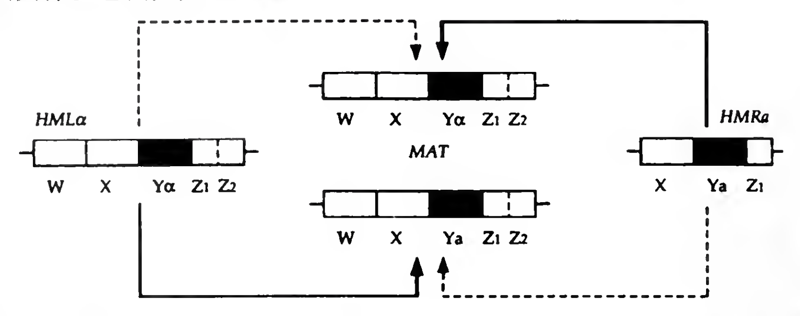


图 25.5 酵母交配型基因座 *MAT* 和沉默信息盒子 *HMLα* 和 *HMRa* 的结构和活性。这三个基因座结构类似，每个都含有一系列的同源序列块，尽管 *HMRa* 比 *HMLα* 短。交配类型是由出现在 *MAT* 基因座的 Y 区域决定。沉默盒子和 *MAT* 之间的重组经常是通过粗箭头表示的途径发生的而产生交配型转换。大约 10% 的重组事件涉及断箭头所示的途径，*MAT* 等位基因被同样的等位基因所取代。重组起始于 Y 和 Z 边界处由 HO 内切酶产生的双链断口。沉默信息盒子的表达和内切酶的剪切都是通过它们被抑制的染色质结构所阻断。

锥虫抗原变异　**抗原转换**描述的是锥虫和包柔氏螺旋体(*Borrelia*)中避免引发它们的脊椎动物宿主免疫反应的保护机制。这个机制涉及覆盖细胞表面蛋白的结构改变，只有一个抗原暴露给宿主。在锥虫中，这是**可变表面糖蛋白**(variable surface glycoprotein, VSG)，而在包柔氏螺旋体中是**可变异主要蛋白**(variable major protein)。转换经常在感染时发生，被抗原识别和破坏的病原体通过转换产生不同性质的抗原发动第二次感染。抗原转换加工可以在感染中发生很多次，产生一系列**变异抗原类型**(variant antigenic types, **VATs**)。

锥虫(*Trypanosoma brucei brucei*)有很多表现 VSGs 的基因和基因片段，很多位于不重要的微染色体上，这样的微染色体似乎除了携带 VSG 基因没有其他功能。基因组中可能有几千个 VSGs 顺序，但只有一百个左右被表达，因此产生的所有潜在抗原类型称为**血清同类群**(serodeme)。在表达的基因中，只有一个是**表达连锁拷贝**或称**表达相关基因**，它在任一时刻都有活性。表达连锁拷贝总在几个**端粒表达位点**中的一个中发现，尽管这些位点的位置本身不如在端粒和内部位置发现的转录沉默的基本拷贝那么重要。抗原转换以两种方式发生：第一可能发生基因转换，包括一个基本拷贝被转移到表达位点。这在机制上类似于酵母交配类型转换(见上)，涉及很多 VSG 基因片段的重组，进一步增加了表面蛋白的潜在变异性。第二，转换可以通过关闭一个表达位点而激活另一个位点达到目的。涉及这一过程的精确机制还不清楚，但控制表达位点活性的区域已经被发现，如同酵母交配类型转换一样，这种机制也涉及染色质结构的改变。

25.4　位点专一性重组

位点专一性重组反应　**位点专一性重组**(site-specific recombination)要求在供体和靶分

子中包含有短的特定 DNA 序列，被特定的识别这些序列的蛋白重组酶所催化。位点专一性重组不同于同源重组在于它需要特定的重组序列，而供体和靶分子间的同源性对重组并非关键。根据重组所涉及的体系和重组位点的方向，位点专一性重组可以介导三类反应：

(1)分子间位点专一性重组产生**整合**或**融合**；

(2)重组位点的正向顺序间的分子内位点专一性重组产生**剪切**和**解离**，是整合和融合的相反过程；

(3)重组位点的反向重复顺序之间的分子间重组产生**倒位**。

位点专一性重组系统分成两类。λ 整合/切离反应(**λ 整合酶家族**，λ integrase family)要求松弛的拓扑学结构，能够进行上述三种反应中的任何一种。λ 整合酶类系统包括酿酒酵母 2μ 质粒的 FLP/FRP 重组酶系统和大肠杆菌基因组中抗生素抗性盒在整合位点的整合(见第 13 章可移动的遗传因子)。Tn3 解离反应(**Tn3 解离酶家族**)要求精确的拓扑学结构，获得的反应产物是精确定义的。Tn3 解离酶类体系包括沙门氏菌属(*Salmonella typhimurium*)和噬菌体 Mu 的宿主范围中控制相变(phase variation)的可倒位基因组片段。了解得较清楚的位点专一性重组系统在文框 25.2 中讨论。

25.5 免疫球蛋白的产生和 T-细胞受体的多样性

抗原受体多样性中重组的作用　　脊椎动物免疫系统产生两种类型的抗原识别蛋白：作为膜结合 B 细胞受体(BCRs)或是作为分泌性抗体的免疫球蛋白(Ig)，以及 T 细胞受体(TCRs)(文框 25.3)。抗原识别蛋白可以根据任何获得的抗原产生，也就是说免疫系统必须识别 10^6 到 10^8 不同分子构型。这种显著的多样性反映了相应基因的基本多样性(**胚系的多样性**)，和单个淋巴细胞通过位点专一性重组来重排基因片段产生新的组合(**体细胞多样性**)的能力。抗原特异性的**独特型多样性**(idiotypic diversity)反映了体细胞重组活性和单个淋巴细胞的特性。

Igs 和 TCRs 证明存在**异型多样性**(allotypic diversity)(物种个体间的多样性，反映了等位基因的变化，也就是多态性)和**同型多样性**(isotypic diversity)(物种所有个体的多样性，通常反映了紧密相关基因的表达或不同的剪接变化)。在免疫球蛋白重链基因座中，新型的同型多样性也是通过位点专一性的重组产生的。这决定了每个 Ig 分子恒定区的结构(也就是说这个区域和效应功能有关而和抗原识别无关)。

通过 V(D)J 重组的独特型多样性　　Ig 和 TCR 的抗原结合专一性(独特型多样性)是由可变区决定的，可变区是通过**体细胞重组**由 V、(D)和 J 片段拼接产生的，也就是说重组在前 B 和前 T 细胞中发生，而不是在种系中发生。在三个高度可变的分子互补性决定区域中(complementarity-determining regions)，CDR1 和 CDR2 由 V 片段决定而 CDR3 显示了更多的可变性，由 V(D)J 拼接。多样性不仅通过特定的 V-、(D)和 J-片段选择产生，而且也通过它们不精确的连接——在编码接头的变化可以产生核苷酸的插入或缺失(表 25.4)。进一步的多样性是通过可变区的随机突变产生的（参见体细胞超突变）。多样性片段中，D-J 先连接，接着 V 片段加到 DJ 拼接块上；连接反应后于并依赖于前重排基因的转录反应。特定重排中使用的片段是随机选择的，尽管对某些组合可能有倾向性。

表 25.4　在抗原－受体基因产生独特型多样性的机制

多样性机制	分子基础
组合多样性	不同 V,(D)和 J 片段的使用
	多 D 片段的使用
连接多样性	外切酶在编码接头处降解,导致基因片段连接处位置的变化
N 区域多样性	通过末端脱氧核苷酸转移酶在编码接头掺入随机核苷酸
V 基因取代	在一个已拼接的但为非生产性的 V(D)J 外显子和剩余的 V 片段之间进一步的重组
体细胞超突变	可变区中的点突变,有助于亲和力成熟;只发生在免疫球蛋白基因座(见突变和选择)

V(D)J 重组的机制　V(D)J 重组是 V-,D-和 J 片段两侧的**重组信号序列**(RSSs)位点专一性的重组反应。RSSs 在 V 片段下游、J 片段上游和 D－片段两侧发现。RSS 含有七聚体回文顺序和富含 AT 的九聚体,它们分别含有保守顺序 CACTGTG 和 GGTTTTTGT。七聚体总是邻近于基因片段,通过一个不保守的大约 12bp 或 23～24bp(分别对应于一个和两个螺旋转角)的片段与九聚体间隔开。重组只发生在有不同间隔的 RSSs 之间,称为 **12－23 原则**。两个 RSSs 在包含 RAG1 和 RAG2 蛋白的联会复合体中排列,导致在每个七聚体和基因片段接头处剪切 DNA(图 25.6)。转酯反应在两个基因片段间形成**编码接**

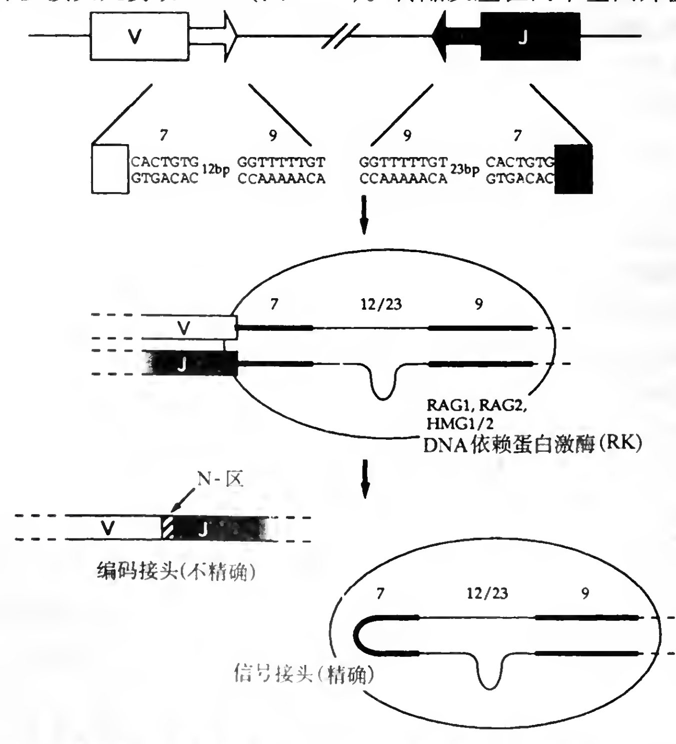


图 25.6　在 IgL 基因座中 VJ 重组机制。有不同大小间隔的重组信号顺序(RSSs)在含有 RAG1、RAG2 和 HMG 蛋白的复合物中联会,使 DNA 在 RSS 和 V 和 J 片段之间剪切。这样形成精确的信号接头(沿残基重组复合物被弃去)和不精确的编码接头(外切酶降解的目标和可插入非模板的 N 核苷酸)。

头并在七聚体模体间形成**信号接头**。信号接头是十分精确的，环化的 DNA 片段从基因组中被剪切出来并降解。相反编码接头是不精确的。可通过不特异的外切酶活性从编码末端去除长至 10 核苷酸或通过末端脱氧核苷酸转移酶加上 10 个核苷酸(这些非模板残基含有 **N－区域**)。

在培养的非淋巴细胞中 RAG1 和 RAG2 蛋白足以提供 V(D)J 重组酶活性，用克隆的 RSSs 在体外可以有效发生 V(D)J 重组。RAG1 有*螺旋-转角-螺旋基序*(参阅)，与 RSS 九聚体顺序结合而不论间隔的大小。如果间隔为 12 碱基，RAG2 被高效征集到 RAG1-RSS 复合物中，这是 12－23 原则的分子基础。RAG 蛋白的结合为 HMG1 和 HMG2 所促进(参见*高泳动蛋白*)，它们可以弯曲 DNA，并可能有助于不同顺式作用位点间的相互作用。剪切后，RAG 和 HMG 蛋白的联会复合体仍保持与信号接头的结合，而编码接头被释放，这可能保护信号接头免受外切酶和末端脱氧转移酶作用。

V(D)J 重组体的解离阶段不是由特定的重组酶进行的，而通过细胞中共有的一般双链断口修复(DSBR)机制进行(见诱变和 DNA 修复)。V(D)J 重组和 DSBR 之间的联系是通过小鼠 SCID 细胞①对 X 射线的敏感性，以及在转染 RAG1 和 RAG2 cDNA 的 SDBR 缺陷细胞中不能进行 V(D)J 重组而确立的。小鼠 *Scid* 基因编码 **DNA 依赖蛋白激酶**多聚体的一种成分，这一复合物被征集到损伤的 DNA 处，这个酶的其他成分由对 X 射线敏感的 *XRCC* 基因编码。

等位基因和同型排斥　单个 B 细胞和 T 细胞是**单特异性的**(monospecific)，也就是说它们产生针对单个抗原特异性的免疫球蛋白或受体。然而，因为淋巴细胞是二倍体，有两个免疫球蛋白轻链基因座，在理论上通过不同多肽链的组合，有可能产生八种类型功能性免疫球蛋白和四种类型的 T 细胞受体。观察到的单特异性反映了**等位基因排斥**(allelic exclusion)(每个基因只表达一个等位基因)，而且在 IgL 基因座中还发生**同种型排斥**(isotypic exclusion)(只表达一个轻链基因)。排斥机制目前还不完全清楚，但它是随机的，可被与 B 或 T 细胞受体结合的抗原所激活。最可能的机制是每个基因座第一次产生的重排使它们合成受体，被抗原结合激活后通过 BCR 或 TCR 信号传导反馈抑制了进一步的 V(D)J 重组。

通过类别转换产生的同种型多样性　免疫球蛋白的效应功能由免疫球蛋白重链(C_H)的恒定区决定。有五种亚型的免疫球蛋白：IgM、IgD、IgG、IgE 和 IgA，IgG 与 IgA 还有不同亚类。B 细胞合成的免疫球蛋白类型由特定的位于 V(D)J 重组外显子下游的 C 基因片段决定。C 基因片段代表的每个类型和亚类型(小鼠中 8 个，人类中 9 个)是串联排列的(图 25.7)。类别转换涉及**转换区域**(switch region，S regions)之间的重组，转换区域位于除了 Cδ 所有 C 基因片段上游。这过程首先缺失内部 DNA，将合适的 C 基因片段连接到 VDJ 外显子中；剪切下来的 DNA 呈环状被释放。重组的 VDJ 基因片段以 Cμ 起始表达，产生免疫球蛋白亚型 IgM。转换重组可以在 Cμ 上游的转换区域和除了 Cδ 外任何 3′C 基因上游的转换区域之间进行(IgM 到 IgD 的类型转换不是通过基因座重排，而是通

---

①SCID 是重度联合免疫缺陷，也就是说细胞调节和体液免疫都有缺陷。SCID 可由很多缺陷引起，如 B 细胞和 T 细胞发育失败。在人类中 30％的 SCID 病例是由 *RAG*1 或 *RAG*2 的基因突变引起的，它们控制早期的 V(D)J 重组。在小鼠中，SCID 的主要形式涉及 DSBR 缺陷，DSBR 涉及晚期 V(D)J 重组，效应不仅仅限于淋巴细胞。

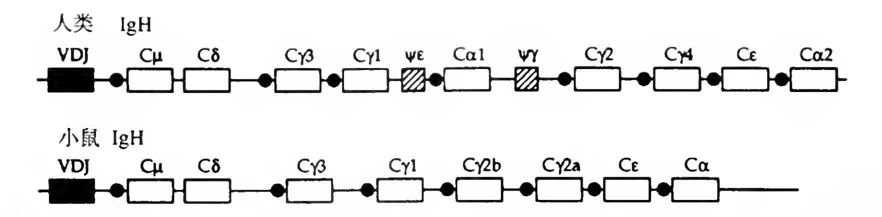


图 25.7 人和小鼠 IgH C 基因片段的组成,黑圈代表转换区域。人类 IgH 基因座含有两个假基因,跨越大约 1.2Mbp DNA。重组在 Cμ 转换和其他 3′端的转换之间进行。接着进一步的转换重组以 5′→3′方向发生。

过选择性的 RNA 加工,引起 Cμ 片段在 RNA 水平去除。在 IgH 基因座的 RNA 加工也用来产生分泌和膜结合的免疫球蛋白亚型)。几个顺序进行的转换重组事件也可能发生,通常是以 5′→3′方向,产生不同亚型的抗体而不影响它们的抗原特异性。转换区域是含有简单串联重复序列的 1～10kbp 的顺序。与 VDJ 重组不同,转换重组可以在转换区域内很多位点发生。转换区域内和上游的很多顺式作用元件被发现可作为蛋白结合位点,尽管已分离鉴定了几个结合蛋白〔包括 NF-κB,NF-Sm(SNUP),LR1 和 Pax-5〕,它们在重组及其调节中的作用还不清楚。转换似乎不是程序化的,但特定的重排的进行能够被细胞因子调节,因此细胞可以左右特异球蛋白类型和亚型的产生。转录在重组之前发生,但顺序剪接对转换的发生很重要。这可能反映了转录时剪接因子将重组蛋白征集到 DNA 上,形成三链的 RNA-DNA 结构或者将 DNA 置于细胞核内正确部位使转换可以发生。转换伴随着转换接头两侧 DNA 的突变,说明有错误倾向的 DNA 合成形成了重组机制的一部分。转换重组因此可能涉及拷贝-选择机制(参阅)。

在其他脊椎动物中免疫球蛋白的多样性　　上文讨论的免疫系统多样性的机制是哺乳动物特有的,特别是在小鼠和人类中。在其他脊椎动物中,甚至是其他哺乳动物所涉及的机制有微妙的差别。如在鸡中,尽管免疫球蛋白基因座类似其他哺乳动物,分成不连续的片段,每个类型片段只有一个特定拷贝是持续表达的,通过涉及 V-,(D)和 J 片段多沉默拷贝的基因转换产生免疫球蛋白的多样性。这些曾被定义为假基因,但从原理上与酵母沉默交配类型盒子和锥虫的 VSG 盒子(见上文)没有区别。在低等脊椎动物中,似乎有可观的预先存在的胚系多样性。在这些动物体内也有 V-,D-和 J 片段,但在很多情况下,似乎在胚系中已经重排,不是已经连接在一起,就是以只有一种重排是可能的方式紧密相联。尽管组合多样性受到限制,但编码的连接是不精确的——如同它在哺乳动物中那样——通过缺失和插入核苷酸可产生可观的连接多样性。

25.6　非常规重组

末端连接,异常复制和重组　　**非常规重组**(illegitimate recombination)描述了混杂的重组机制,重组伙伴之间需要很少甚或不需要同源性,通常代表了使用不正确底物的正常细胞加工。非常规重组经常导致基因破坏而引起突变,在人类遗传疾病和癌症中占主要地位。在表 25.5 中总结了非常规重组的部分机制。

表 25.5 非常规重组的机制

机制	分子基础
不正常末端连接	在真核细胞中经常发生的反应，通过直接重接末端来修复双链染色体断裂。通过连接修复被认为是先进的，因为在高等真核生物的高度重复基因组中进行同源寻找是困难的且有错误倾向。当加工确实出错时，丰富的非编码 DNA 降低了基因被破坏的危险，另外因为当 DNA 包装到染色质时，打断的染色体末端可能通过高级结构保持在一起。当修复系统可以作用于不正常末端时，可能会引起易位和其他重排。这种机制还允许线性 DNA 转染并整合到基因组中(参见**转染**、**转基因小鼠**)。末端连接在原核生物中较少，在原核生物中不能连接没有粘端的 DNA。
非常规复制	这发生在重复 DNA 和形成稳定二级结构(发夹、十字形)的 DNA。引物链从带有模板的固定位置中跳出并产生插入和缺失。能刺激产生不正常复制的重复顺序和自身互补的基序经常是突变热点(参见**复制的保真度**、**三核苷酸重复综合征**)。
非常规链交换	所有三种正常重组－同源、位点专一性和转座——产生作为反应中间体的单链。当这些游离链连接到错误的位置，不正常的链交换就会发生，如一单链正好位于反应附近的情况。这也可能在通常不涉及重组的系统中发生，如**拓扑异构酶**(参阅)和起始质粒复制的**缺口酶**(参阅)。
不等交换	如果由于存在重复 DNA，同源双链突出固定位置。解离之后，它在一条链产生插入，而在另一条链产生缺失，从而发生不等交换。尽管不等交换是异常的，但机制是正常的同源重组而不是不正常的重组——它的发生需要同源性。
异常位点识别	重组基于特定序列被蛋白识别时邻近 DNA 出现类似序列(隐蔽识别位点)有时造成涉及额外 DNA 的不正常反应。这种类型的重组导致大多数涉及转座和位点专一性附加体的不正常切离事件。
非常规 V-(D)-J 连接	免疫球蛋白的 RSS 元件和 TCR 基因座和在基因组别处的隐蔽元件间的不正常重组可导致与淋巴瘤相关的染色体畸变谱型。另外，在正常情况下在 B 和 T 细胞发育中表达的重组酶基因若在异位激活会产生肿瘤。隐蔽的 RSS 元件在抗原受体基因座本身可以发现。**Kappa 缺失元件**就是这样的隐蔽位点，它涉及的重组引起 IgLκ 恒定区的缺失，在小鼠中称为 **RS 重组**的过程。类似的 **delta 缺失元件**引起 TCRδ 链的丢失。在 SCID 小鼠中，当 DJ 连接阻断时，外切酶活性异常，在 IgH、TCRβ 和 TCRγ 的 DJ 编码接头两侧发生缺失，结果不能产生成熟的 B 和 T 细胞。

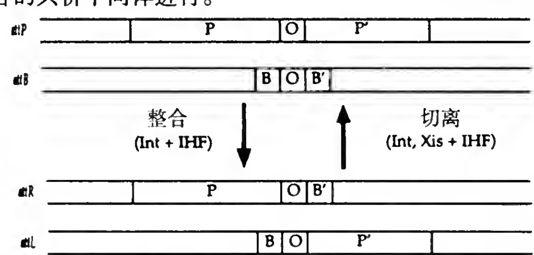
文框 25.1 负责同源重组的基因

大肠杆菌中的重组蛋白：RecA 是多功能蛋白，帮助联会和链交换，在 SOS 修复中也起主要作用。*recA* 突变体不能进行重组，也不能诱导 **SOS 反应**(参阅)，因此对 DNA 损伤试剂非常敏感。RecA 可以和单链或双链 DNA 形成复合物，而其中单链 DNA 是更有效的底物。RecA 沿 DNA 多聚化形成螺旋状，**突触丝**(presynaptic filaments)中的 DNA 较大程度地解盘旋(比正常长 50%)。RecA-DNA 复合物与双链 DNA 相互作用，形成四元复合物，如果发现有同源区域，就起始链转移。四元复合物也可以在没有顺序相似性的情况下形成，这被认为反映了通过蛋白寻找同源性的恶果。蛋白丝有明确的极性，与 5′→3′的进入和取代链的极性相一致。蛋白促进单链和同源双链中的互补链的碱基配对，首先通过形成含有 RecA 和重组对的四元复合物，然后作为尾或环取代受体双链原先的一条链。这个反应被认为涉及三链 DNA 和可能的异常碱基配对(见**核酸结构**)。进入和取代链的链进入和取代以 5′→3′方向进行。起先，配对必须是完全的，但一旦链开始转移，错配可以被忍受，导致异源双链 DNA 的产生。一旦链转移开始，在起始复合物中 RecA 蛋白可以催化取代链转移到不配对的链，因此产生霍利迪接头。RuvAB 蛋白在 RecA 链延伸的方向促进分枝迁移，而 RecG 促进分枝向相反方向迁移。RuvC 剪切霍利迪接头，以相同频率产生短片段型和拼接型重组体(参见 **D-环诱变**)。

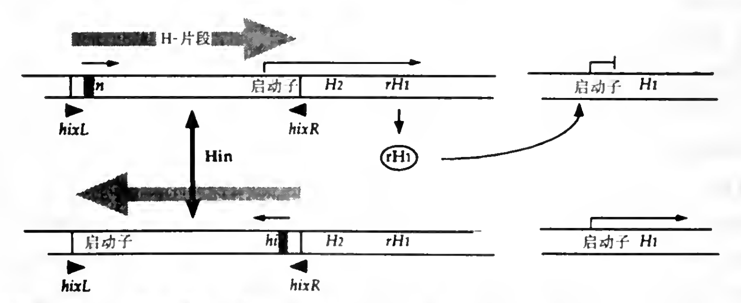
真核重组蛋白　分离酵母 DNA 损伤修复和减数分裂阻断突变体可促进重组基因的鉴定。RAD52 上位性群(参阅)基因涉及重组修复。这些基因中占优势的是 RAD51,它与大肠杆菌 recA 同源。RAD51 的同源物在很多真核生物中被发现,酿酒酵母的 RAD51 被 DNA 损伤诱导。在每个物种中似乎都有几个 RAD51 的同源物。如在酿酒酵母中,DST1 编码类 RAD51 产物,它的活性限于减数分裂细胞。其他基因有突变体表型,表明它们在减数分裂和有丝分裂重组中都起一定的作用。RAD50 编码一 DNA 结合蛋白,它的功能是在有丝分裂细胞中修复双链断裂点,并在减数分裂细胞中诱导和加工这些断裂点。RAD52 的产物对有丝分裂和减数分裂重组,双链 DNA 断裂点的修复和交配类型的转换都很重要;它与 RAD51 相互作用。RAD54 编码一解旋酶,为重组和其他形式的 DNA 修复所必需。

文框 25.2　位点专一性重组系统

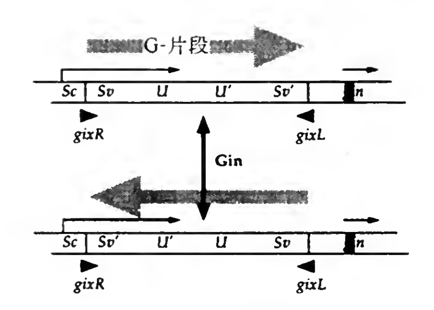
λ 噬菌体的整合和切离　在整合和切离中需要两个 λ 编码的蛋白,Int 和 Xis (关于它们在 λ 感染循环中的表达和调节讨论见病毒一章)。它们与宿主的 IHF 和 FIS 联合调节称为**附着位点**(attachment site)序列间的位点专一性重组,这些位点是噬菌体附着位点 *attP* 和细菌附着位点 *attB*。整合需要 Int 和 IHF,而切离还需要加上 Xis(FIS 有促进作用,但对切离不是关键的)。链交换反应是按顺序进行的,首先是在附着区域的左侧接着通过分枝迁移到第二链转移发生的右侧。这些反应通过一 DNA 与 Int 结合的共价中间体进行。



在噬菌体 λ 整合和切离中的位点专一性重组。噬菌体附着位点 *attP* 有 POP′结构,P 和 P′是含有 Int、Xis 和 FIS 结合位点的复杂元件,而 O 是位于链转移位点间的重叠区域。细菌附着位点 *attB* 简单些,有结构 BOB′,其中 B 和 B′是 Int 的结合位点。位点专一性重组产生在 λ 原噬菌体两侧的杂合位点 *attL*(BOP′)和 *attR*(POB′)。



沙门氏菌的**相转变**。沙门氏菌的鞭毛主要由蛋白鞭毛蛋白组成,这种蛋白有两个不同的抗原形式,*H1*(特殊相位抗原)和 *H2*(群相位抗原)。*H1* 和 *H2* 由不共表达的不同基因编码,因为 *H2* 基因座与第三个基因 *rH1* 紧密连锁,*rH1* 编码 *H1* 表达的抑制因子。*H2* 和 *rH1* 的启动子位于 *hixL* 和 *hixR* 位点间的倒位 **H**-片段中,可在 Hin 的控制下进行位点专一性倒位。在一个方向上,启动子驱动 *H2* 和 *rH1* 的表达,抑制 *H1*。在另一方向,*H2* 和 *rH1* 缺乏启动子,因此 *H2* 不表达而 *H1* 表达去抑制。宿主蛋白 FIS 通过与 *hin* 基因中的增强子结合促进反转。



噬菌体 Mu 的宿主范围变化。在噬菌体 Mu 中有两种尾丝基因(*S* 和 *U*)。*S* 基因有一个位于可倒位 **G 区域外**侧的不变部分(*Sc*)和位于其中的不同的可变部分(*Sv*、*S′v*)。不同版本的 *U*(*U*、*U′*)也位于 G 区域中。G 区域的倒位由蛋白 Gin 催化,Gin 帮助两侧 *gixR* 和 *gixL* 位点间的重组。宿主蛋白 FIS 通过与 G 区域外侧 *gin* 基因中的增强子结合刺激倒位。

通过位点专一性倒位发生表型变化　在有些系统中,如在沙门氏菌中的相转变和噬菌体 Mu 的宿主范围的变化,不同的基因表达模式依赖于小的遗传元件的倒位。这样的倒位因 Tn3 解离酶类型的位点专一性重组系统而促进,利用可倒位元件两侧的反转重复顺序进行。相转变和 Mu 宿主范围变化的机制在下图中显示,类似的系统控制噬菌体 P1 和大肠杆菌基因组中未知功能的小 DNA 片段倒位。涉及这些系统的重组酶,Gin、Cin、Hin 和 Pin 是彼此同源且功能可以互换的,但涉及倒位的顺式作用元件不是这样。在噬菌体 Mu 和 P1 的 Gin/Cin 系统中,重组酶的基因也含有倒位的顺式作用增强子,并位于可倒位片段的范围之外。在沙门氏菌的 Hin 系统中,重组酶和增强子在倒位片段内而涉及的变异基因位于片段之外。

通过位点专一性重组解离　转座子 Tn*3* 从一个基因组到另一个发生复制转座之后,产生含有一个转座子的正向重复顺序的共整合(参见复制转座,共整合)。在转座因子内部 *res* 位点发生位点专一性重组,来共整合两个各携带一个 Tn*3* 拷贝的 DNA 片段。这个重组反应由转座子编码的**解离酶**(resolvase)蛋白催化。与 λ 的整合和切离不同,解离酶的活性是有方向性的。融合(整合)反应不是由这个酶催化的,反映了系统的特定的拓扑学需要。这表明 Tn*3* 在体外分解经常产生同样拓扑学状态反应产物。反应经历两条链交换反应,涉及共价结合的 DNA-解离酶中间体。另一个控制解离的位点专一性重组是噬菌体 P1 的 **Cre-*loxP* 系统**。它完全不依赖于上面描述的 Cin-*cixL*/*R* 系统,它的基本功能是解离质粒二聚体,以协助细胞分裂时的精确分离(噬菌体 P1 以低拷贝数质粒形式存在,在处于溶原状态时以 F 质粒同样方式繁殖;见质粒、病毒)。Cre 重组酶作用于正向重复的 *loxP* 重组位点,帮助解离,但也能催化线性噬菌体基因组的环化和(很罕见)可能将质粒整合到宿主基因组中。酵母 2μ 质粒的 FLP-FRP 系统也被认为在分解质粒多聚体中发挥功能,也可能在增加拷贝数中起作用。与 Tn*3* 解离酶和 Cre 不同,FLP 重组酶属于 λ 整合酶家族,除了协助解离还催化倒位和融合。在这些系统中,FLP-FRP 和 Cre-loxP 已经被广泛用作基因组操作和基因调节的工具(参见转基因动物、剔除小鼠)。

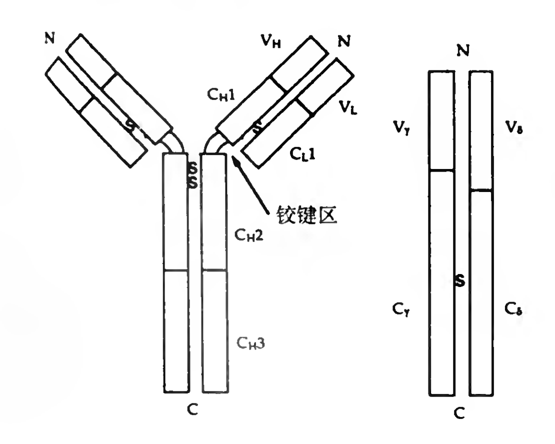
文框 25.3 抗原受体和它们的基因

蛋白 **免疫球蛋白**和**T细胞受体**分别是由B淋巴细胞和T淋巴细胞合成的抗原结合蛋白。免疫球蛋白可以作为**B细胞受体**出现在细胞表面或作为**抗体分泌**:两种分子都有基本的异源二聚体结构,两个相同的**重链**和两条相同的**轻链**,由二硫键稳定。只有一种重链类型(IgH),而存在两种轻链类型(IgL κ和IgL λ),其中任何一种都可以组成免疫球蛋白分子。有四种TCR单体(α β γ和δ),这些形成两种类型的异源二聚体(α β-TCR和γ δ-TCR)。

免疫球蛋白和TCR链有类似的组织构成,一个负责抗原结合的N端**可变区**(variable region)和负责效应功能的C端**恒定区**(constant region)(这些对所有给定亚型都是共同的,无关抗原的专一性)。一个免疫球蛋白的整个结构是Y型的,这是因为重链的铰链区域:Y形叉含有重链和轻链的可变区和与抗原结合的钳子结构。可变区包括中等可变的**框架区域**(framework regions)和**超变区域**(hypervariable regions)或者**互补决定区域**(complementarity determining regions, **CDRs**),前者以正确的三维构象与后者联系,直接与抗原相互作用。有些不同类型和亚类型的免疫球蛋白在免疫系统中有不同的功能,抗体类型是由不变区的结构决定的。不同的恒定区也可以在TCR β和TCR γ分子中发现。下图反映了一般免疫球蛋白和γ δ-TCR的结构,标明了恒定区和可变区的位置。

基因 哺乳动物免疫球蛋白和TCR基因是唯一的在体细胞中将由胚系提供的一系列可选择性的不连续的"建筑块"组合而成。所谓的**胚系构型**(germline configuration)是非功能的,基因重排对产生功能的受体和抗体是关键的。

每个分子的恒定区是位于每个基因座3′末端的C片段决定的(在免疫球蛋白基因座,C片段串联排列代表了每个亚型的恒定区)。IgH和TCR β和γ基因的可变区是由三种类型的基因片段组建的——**V**(可变)**片段**,**D**(多样性)**片段**和**J**(连接)**片段**。V-,D-和J-片段排列使不同结构的可变区可以通过**重组连接**产生。IgL和T细胞α和δ基因的可变区仅由V和J片段拼接产生。各片段的性质和数目在不同基因座不同,从小鼠和人类基因组得到的片段在下表中显示。一般,免疫球蛋白基因相对结构简单,是不同的片段以同样的顺序拼接。T细胞受体有更复杂的结构。*TCRD*基因处在*TCRA*基因中,V片段混合排列。另外,V片段位于对应C基因片段的上游和下游,涉及下游V基因的重排会引起大的反转。



种类	片段的基因座和数目				评论
	V	D	J	C	
人类					
IgH	86	～30	9	9(+2ψ)	
IgL κ	76	–	5	1	
IgL λ	52	–	7	7	JC 重组预先决定
小鼠					
IgH	＞1500	12	4	8	
IgL κ	＞200	–	4(+1ψ)		1
IgL λ	2	–	4	4	JC 重组预先决定

（陈　淳 译）

第26章
复　　制

基本概念和定义

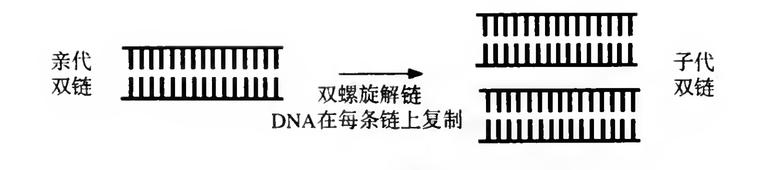
- **复制**(replication)可广义地定义为基因组复制,是细胞基因组和那些"分子寄生虫"——病毒、质粒和转座因子繁殖的重要步骤。被复制的基因组称为亲本基因组,而复制产物是**子代**基因组。
- 所有细胞基因组都是双链DNA。然而有单链DNA和RNA基因组的病毒和其他遗传元件也必须被复制,传染性海绵状脑病(参阅)的传播媒介被认为是在蛋白水平上的一种复制。
- 在生物化学水平上,复制被定义为模板指导的核酸合成反应,模板和新生链是相同类型的核酸。这与一条链是DNA而另一条链是RNA的转录和反转录不同(文框26.1)。双链DNA的复制以**亲本双链**分成两条链为开始,每条作为模板在一步加工过程中产生两个一样的**子代双链**,称为**直接复制**(direct replication,图26.1)。两个子代双链互为**姐妹双链**(因此象姐妹染色单体交换一样称呼)。然而,因为每个新合成的链是与亲本链互补,而不是与其完全一样,单链基因组的复制必须通过两步加工的过程,第一个合成反应产生与亲本链反义的中间体,中间体本身可以作为模板产生亲本链的复制分子。这样称为**非直接复制**,可以涉及任何四种核苷酸合成反应(图26.1)。
- 复制是一个多聚化反应,可以分成起始、延伸和终止阶段。延伸的复制中心需要很多不同酶活性的协同作用,总称为**复制体**。

26.1　复制的策略

DNA复制的模型　　细胞DNA的复制起先被认为有两种模式:全保留和分散模型。根据**全保留复制模式**,亲本DNA保持不变,进入一个子代细胞,而新合成的DNA进入另一个细胞。根据分散复制模式,新合成的DNA和亲代DNA相间排列,每条子代DNA链[1]接受了亲本和新合成DNA的混合物。

第三种机制**半保留复制**模型是由James Watson和Francis Crick根据他们所阐明的双螺旋DNA结构提出的,亲本链保持不变,但双链分成两半。每个亲本链作为模板复制,子代双链包含一条亲本链和一条子代链。半保留复制模型适用于细胞DNA,但病毒

1)此处应为DNA方合理,原文误写成细胞(cell)。——译者注



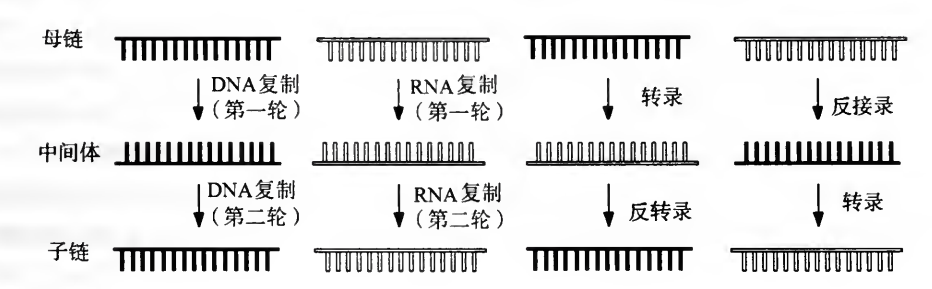


图 26.1 单链和双链基因组的复制。双链 DNA 基因组的复制涉及一轮 DNA 复制，不需要中间体。单链基因组的复制涉及任何类型的核酸合成反应，因为它涉及中间体。黑线是 DNA，白线是 RNA。

和一些质粒的单链基因组是全保留复制的——亲本链的结构通过复制保持不变。〔Meselson-Stahl 实验：在 1958 年，Matthew Meselson 和 Franklin Stahl 证明细菌染色体 DNA 的复制是半保留的。大肠杆菌在含^{15}N 的培养基中生长，所以它们的 DNA 被广泛标记上同位素(重 DNA)。然后细胞被转移到含正常^{14}N 的培养基中，DNA 从经过一到两轮复制的细胞中分离出来。DNA 用浮力密度离心分析，分辨重链 DNA，正常的轻链 DNA 和包含一条重链和一条轻链的中间体 DNA。在一轮复制之后，DNA 都是处于中间体的密度中，而在两轮复制后，有等量 DNA 处于中间体和轻链密度。这些结果只与半保留复制模型相符。所有细胞基因组都是半保留复制的。在真核生物中，半保留复制可以通过将溴脱氧尿苷掺入到染色体 DNA 中来证明。两轮复制的掺入可以根据一半 DNA 姐妹染色单体两条链含溴脱氧尿苷，而其他只在一条链含溴脱氧尿苷来分辨。染色单体的染色是不同的（参见**斑条染色**）〕。

半保留复制　Watson 和 Crick 的 DNA 复制半保留模型预见有复制叉的存在，复制叉是由亲本双链 DNA 构成桶状和子代双链构成手臂的动态的 Y 型结构，每个子代双链由一个亲本链和一个子代链组成(图 26.2)。在复制叉的中心，亲本双链会解链，核酸加到生长中的子代链上。复制叉的存在通过在复制的细菌 DNA 中掺入放射性核苷酸和在电子显微镜下观察到中间体的结构得到直接证明。然而这个模型提出了以下的明确结论：

(1) 细胞 DNA 是半保留复制的；

(2) 两条子代链是同时延伸的；

(3) 亲本双链是反平行的；

(4) DNA 聚合酶只以 5′到 3′方向延伸 DNA。

当亲代模板有相反的极性，两条子链在复制叉如何同时促进 5′→3′的延伸呢？这是通过**半保留复制**完成的，其中一条链连续延伸，而另一条是短片段的集合而不连续合成。

DNA复制的半保留机制被表示成**先导链-后滞链模式**(leading strand-lagging strand model)。**先导链**(leading strand)是在复制叉移动方向连续合成的新生链,因为它的3′末端暴露于与DNA聚合酶。先导链的模板因此是**正向模板**(forward template)。**后滞链**(lagging strand)是另一条新生链,因为它的5′末端方向不能延伸,因此是与复制叉相反方向进行不连续合成的。后滞链的模板因此是**逆向模板**(retrograde template)。整个机制可以总结如下:当复制叉向前移动,先导链延伸,一部分逆向模板露出来。DNA聚合酶在复制叉前进的反方向移动,合成一小片段DNA,称为**冈崎片段**(Okazaki fragment)。之所以称为后滞链是因为先导链必须先合成以使后滞链模板对应部分暴露出来。当多聚酶到达前面一个已经合成好的冈崎片段时它从模板上解离下来,而这时另一部分的逆向模板已经暴露出来。这样酶就重新起始合成一新的冈崎片段。通过重复这样的"回针"过程,后滞链在表面上是以3′→5′方向合成的(图26.2)。因为DNA聚合酶不能从头起始链的合成,每个冈崎片段需要独立引发。解离-重新结合的循环因此需要引发(参阅)的步骤。

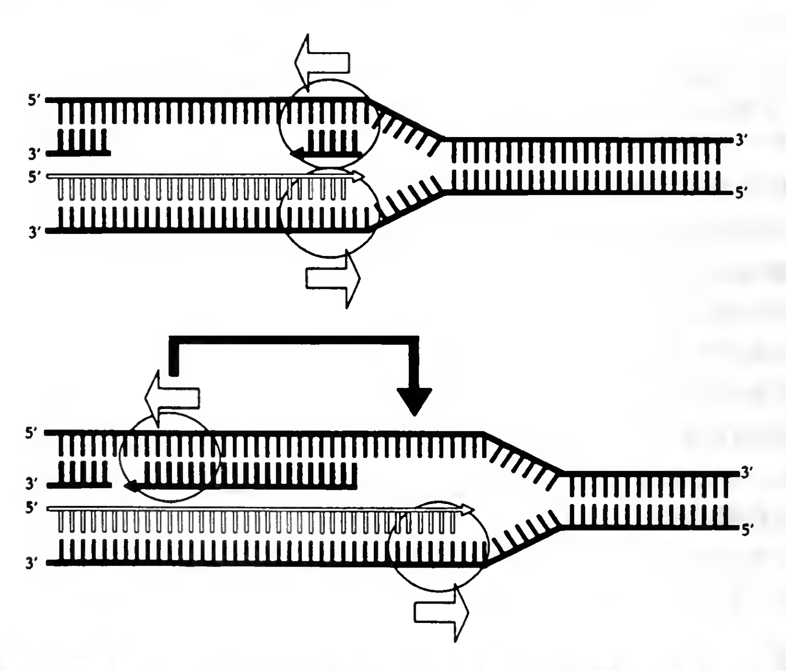


图26.2 在复制叉处DNA半保留合成的模型。复制体包含在同一位置起始的两个DNA聚合酶。当双链解链,先导链的合成是以复制叉移动方向延续的。这样露出一部分后滞链模板,第二个DNA聚合酶以复制叉移动的相反方向合成一短的冈崎片段。当其到达前面的冈崎片段时,酶就解离而到其他模板上未被覆盖的起始位点。值得注意的是尽管两个酶的相对方向是不同的,但它们不一定在物理上是分离的。酶周边的反向模板成环可获得这样的结果。黑线是亲本DNA,白线是先导链,灰线是后滞链。圆环代表DNA聚合酶(合成和多聚酶的移动的方向以小箭头和大箭头分别表示)。为了清晰,RNA引物没有表示出来。

支持先导链-后滞链模型的证据包括细菌和真核生物复制叉处复制体中引发酶的存在。同时,复制体都是不对称的,反映了存在一先导链的高度推进的DNA聚合酶和一后滞链分步性的DNA聚合酶。脉冲追踪实验确认细胞复制中新生DNA的50%是低分子量的片段。

取代复制　细胞基因组和很多 DNA 病毒和质粒的复制是使用上述的半保留复制的机制，它们子代两条链同时延伸，需要复制叉。另一种双链 DNA 复制子使用不同的策略，起始只有一条链作为模板，新生链连续延伸取代另一条亲本链。这是**取代复制**(displacement replication)，代表连续合成新生链的位置的动态结构称为**取代复制叉**(displacement fork)。在很多这样的复制子中(如质粒 ColE1)，被取代的链最终作为不连续合成的模板，而整个机制与半保留复制一样。而在其他一些情况中(如线粒体 DNA 和腺病毒基因组)，先导链的连续合成暴露了被取代的亲本链的第二个起始点，有利于在相反方向的连续合成。在这种情况下不需要冈崎片段，没有后滞链；这样的过程定义为**连续复制**(continuous replication)。

一种特殊类型的取代复制是通过在一环状双链基因组中产生缺口，在自由的 3′末端延伸。在这种情况下，链的延伸将取代原先的亲本链，通过循环多次利用模板连续取代，推出多聚的产物，它作为第二链合成的模板。这种**滚环复制**(rolling-circle replication)的策略在很多噬菌体和格兰氏阳性细菌中占优势。

单链基因组的复制　单链 DNA 和 RNA 基因组的复制经常涉及基因组反义[有时称为**反基因组**(antigenome)]的**复制中间体**。复制中间体可以作为子代基因组合成的直接模板，或者是基因组和反基因组相互作用形成双链的**复制型**(replicative form)，从而子代基因组通过取代复制产生。在逆转录病毒的特殊例子中，一旦 DNA 反基因组合成完成，亲本 RNA 基因组被破坏，被第二条 DNA 链所取代 - 子代基因组通过这种双链原病毒的转录产生。进一步的讨论见病毒。

26.2　细胞复制体和延伸的酶学

复制体　**复制体**(replisome，或称**复制复合物，中心机器**等)是延伸的复制叉中酶和其他蛋白的动态复合物。尽管复制机制非常简单，连续精确的 DNA 合成需要很多酶活性协同作用。超螺旋 DNA 解旋需要极大的能量，在真核生物中，DNA 组织成染色质必须与复制体相协调(见第 3 章染色质)。细胞复制体的成分和功能在表 26.1 中列出，细菌和真核生物复制体成分的比较在表 26.10 中。在起始时，复制体由它的成分组建，只在复制叉中发现其存在——它在细胞中不作为一个独立的实体存在。简单的复制子可能需要少量的复制体成分，最低的要求是一个简单的聚合酶。

复制缺陷突变　原核和真核复制体系统成分的阐明依赖于 DNA 复制缺陷突变体的分离。复制是一非常重要的过程，所以相关蛋白经常是通过温度敏感的条件突变(参阅)来鉴定。在细菌中，复制突变分成两类：快停和慢停。**快停突变**(quick-stop mutant)当转移到不允许的温度时，不能完成目前一轮的复制，可鉴定得到对延伸重要的基因。**慢停突变**(slow-stop mutant)能完成这轮复制，但不能再起始下一轮复制，可鉴定起始或终止相关的基因。有些基因同时是快停和慢停的等位基因，因此可能在多个阶段起作用。有些是突变子等位基因(参阅)；它们影响 DNA 复制的精确性或 DNA 修复(参阅)。值得一提的是很多复制缺陷突变体在 DNA 代谢的其他方面也有缺陷(重组、修复、转录)，表明有些基因在 DNA 加工中有普遍的作用。表 26.2 提供了鉴定复制体成分的一系列 *E. coli* 复制

突变体。

表 26.1　细胞复制体的遗传成分

复制体成分	在复制中的功能
DNA 解旋酶	使复制叉前面的 DNA 解链
DNA 连接酶	连接已修复的后滞链片段
DNA 聚合酶	DNA 合成,后滞链缺口的修复。复制体可能含有几种不同形式的 DNA 聚合酶,一种负责先导链的合成,另一种负责后滞链的合成,还有一种负责后滞链的修复
DNA 引发酶	引发冈崎片段的合成
DNA 拓扑异构酶	释放因解旋酶的活性而造成的扭曲的链。复制后使连接的环去链接(参见 DNA 拓扑学)
RNaseH	从后滞链中去除 RNA 引物
单链结合蛋白	稳定复制叉的单链区域。可能与复制体其他成分相互作用促进其活性

拓扑异构酶在复制叉处没有被发现,但在扭曲浓缩 DNA 的复制中是不可缺少的。见表 26.10 中细菌和真核生物复制体的特殊成分。

表 26.2　一些 *E. coli* 复制缺陷突变体和每个基因座的功能

基因座	表型	功能
dnaA	慢停	起始蛋白
dnaB	快停	解旋酶
dnaC	快停和慢停等位基因	与 DnaB 解旋酶形成复合物,辅助起始位点的定位
DnaE(*polC*)	快停	多聚酶 III 全酶的 α 亚基
dnaG	快停	引发酶
dnaN	快停	多聚酶 III 全酶的 β 亚基
dnaQ	快停和突变子等位基因	多聚酶 III 全酶的 ε 亚基
dnaT	慢停	引发体成分
danX	快停	多聚酶 III 全酶的 γ 和 τ 亚基
gyrA	快停与慢停等位基因	DNA 回旋酶亚基 α
gyrB	快停与慢停等位基因	DNA 回旋酶亚基 β
lig	修复缺陷	DNA 连接酶
ori	致死(不是条件性)	复制起点
polA	修复缺陷	DNA 聚合酶 I
rnhA	稳定复制	RNase H
ssb	快停	单链结合蛋白
ter	无	复制终止
tus	无	终止蛋白

此表列出了那些在复制叉或起始、终止中起作用的顺式作用元件和基因。影响 DNA 代谢或噬菌体或病毒基因组复制能力的基因没有列出。

DNA 聚合酶　在 DNA 模板上催化 DNA 合成的酶是 **DNA 聚合酶**。它们在细胞中行使两个基本功能:基因组复制时的 DNA 合成和损伤或重组后(以及后滞链引物切除后)缺少的 DNA 的重新合成。在原核和真核生物中,特定的 DNA 聚合酶致力于复制和修复功能,前者有时称为 **DNA 复制酶**(DNA replicases)。所有的 DNA 聚合酶都有 **5′→3′多聚酶活性和焦磷酸化酶活性**,都促进 DNA 合成。与 RNA 聚合酶不同的是,DNA 聚合酶不能

从头起始链的合成，需要引物。大多数 DNA 聚合酶也有进一步的内在活性(表 26.3)。最重要的是 3′→5′**外切酶活性**，是校正功能的基础(见诱变和 DNA 修复中复制保真度的有关讨论)。

大肠杆菌 DNA 聚合酶　大肠杆菌基因组编码三个 DNA 聚合酶(**DNA 聚合酶Ⅰ，Ⅱ**和**Ⅲ**，或者表示为 **pol Ⅰ，pol Ⅱ**和 **pol Ⅲ**)。这些酶的性质在表 26.4 中总结。pol Ⅰ和 pol Ⅱ是单链多肽，其基本功能似乎与 DNA 修复有关。pol Ⅰ(也称为 **Kornberg 多聚酶**)是细胞中主要的聚合酶，有独特的 5′→3′外切酶活性，这种活性可以在体内修复合成中帮助从后滞链切除引物 (参见缺口平移)。尽管所有的 pol I 酶活性存在于 109kDa 的单一多肽链中，每一种酶活性都是从特定的结构域中产生的，蛋白酶剪切可以产生一个大的 C 末端片段(**Klenow 片段，Klenow 多聚酶**)，它缺乏 5′→3′外切酶活性，在体外不希望剪切的实验操作中非常有用(参见随机标记、体外突变)。Klenow 聚合酶是通过表达截断的 *pol A* 基因生产的。

表 26.3　DNA 聚合酶相关的内在酶活性

DNA 聚合酶活性	功能	DNA 聚合酶活性	功能
所有 DNA 聚合酶的活性		有些 DNA 聚合酶相关活性	
模板结合活性	底物识别	3′→5′外切酶活性	校正
核苷酸结合活性	底物识别	5′→3′外切酶活性	引物切除和修复
5′→3′多聚酶活性	DNA 合成	引发酶活性	引物合成
焦磷酸酶活性	DNA 合成		

表 26.4　大肠杆菌 DNA 聚合酶性质的比较

	聚合酶Ⅰ	聚合酶Ⅱ	聚合酶Ⅲ
5′→3′多聚酶	√	√	√
3′→5′外切酶	√	√	√
5′→3′外切酶	√	×	×
结构	多肽	多肽	多亚基复合物
大肠杆菌基因	*polA*	*polB*	*dnaE*，*dnaN*，*dnaQ*，*dnaX* 和其他未知亚基
功能	基本的修复聚合酶和引物切除	错误倾向修复聚合酶(SOS 诱导)	基本的复制聚合酶

pol Ⅱ是细胞正常生长中一种次要成分，被 SOS 反应(参阅)所诱导。它似乎允许核苷酸掺入 AP 位点的对面(参阅)，也就是说，阻断 pol Ⅰ和 pol Ⅲ的损伤处——因此有利于反损伤合成(translesion synthesis，参阅)。

大肠杆菌的基本 DNA 聚合酶是 pol Ⅲ，它与其他酶不同，是多亚基复合物，称为 **pol Ⅲ全酶**(表 26.5)。全酶在复制叉上以异源复合体发挥功能，每个单体看来与一子代链的合成有关。在体外，α、ε 和 θ 亚基组成核心酶，包含主要的酶活性。其他亚基的添加促进二聚化和增加了酶活性。体内全酶的组建是这样的：β 亚基以二聚体发挥功能，形成一个环或夹可以沿单链 DNA 滑动。这是推移因子，使核心酶附着在模板上。β 亚基通过 γ 复合物定位在模板-引物上，这是一个依赖ATP的过程，形成**起始前复合物**。β亚基的定位使

表 26.5　大肠杆菌 pol III 全酶的亚基和推测的功能

亚基	基因	性质和推测的功能
核心亚基		
α	*dnaE*(*pol C*)	5′→3′多聚酶活性,DNA 合成所需
ε	*dnaQ*(*mutD*)	3′→5′外切酶活性,校正所需
θ	未鉴定	功能不确定,可能帮助其他亚基的组建
附属亚基		
τ	*dnaX*[a]	DNA 依赖的 ATP 酶,起始所需,促进二聚化
γ	*dnaX*[a]	与四个多肽(见下)相结合形成 DNA 依赖的 ATP 酶,称为 γ 复合物,起始所需,帮助 β 的结合
δ,δ′,χ,ψ	未鉴定	与 γ 结合形成 γ 复合物(见上)
β	*dnaN*	增加全酶进行性的"滑动夹"。β 与 DNA 结合形成起始前复合物,这一过程需要依赖 ATP 的 γ 复合物的活性,*dnaN* 被 SOS 反应(参阅)所诱导

核心酶能够结合,τ 亚基的添加有利于二聚化。完整全酶的结构在图 26.3 中表示。全酶除了 γ 复合物外是对称的,γ 复合物只与一个单体结合。β 亚基在 DNA 上定位和解离都需要 γ 复合物,因此 γ 复合物控制了酶的进行性。当聚合酶遇到逆向模板上前面合成的冈崎片段时,γ 复合物的存在允许 β 亚基从模板引物上解离。这样当核心酶完成了冈崎片段,将从后滞链的全酶中释放出来,与下一个由 DNA 引发酶提供的模板引物的起始前复合物结合。

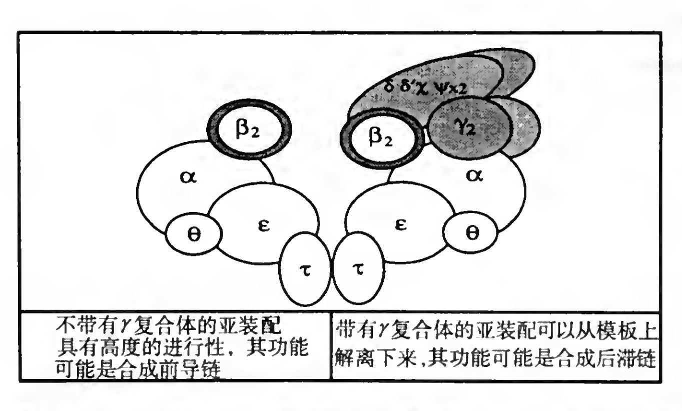


图 26.3　大肠杆菌 pol III 异源二聚体的组建刺激先导链和后滞链的合成。

其他细菌的 DNA 聚合酶类似于大肠杆菌的聚合酶,事实上,所有的 DNA 聚合酶可以根据六个保守结构域的结构分类。最值得注意的多聚酶是嗜热的和嗜温热的原核生物,它们可以在高于 100℃ 时催化合成 DNA,大大超过 DNA 的 T_m 值。在这样的生物中,DNA 的解链可能是被动发生的,但聚合酶必须有将引物固定在模板上的能力以延伸它。*Taq* DNA 聚合酶起先是从栖热菌属(*Thermus aquaticus*)中提取得到的,被广泛用于聚合酶链式反应(参阅),它通过在大肠杆菌中的过量表达进行商业化生产。它与大肠杆菌

pol I 同源,但它缺乏 3′→5′外切酶活性,不能进行校正。其他耐温的 DNA 聚合酶有校正活性(如从 *Pyrococcus furosus* 得到的 Pfu 多聚酶)。

真核生物 DNA 聚合酶　　真核细胞含有四种细胞核 DNA 聚合酶和第五种负责细胞器基因组复制的多聚酶(表 26.6)。细胞核酶是 **DNA 聚合酶** α、β、δ 和 ε。DNA 聚合酶 α 和 δ 负责染色体复制。DNA 聚合酶 α 有相关的引发酶活性,但没有 3′→5′外切酶活性,而 DNA 聚合酶 δ 有校正功能。DNA 聚合酶 δ 与称为**增殖细胞核抗原**(proliferating cell nuclear antigen, PCNA)的辅助因子结合,PCNA 是与大肠杆菌聚合酶 III 的 β 亚基类似的细胞周期依赖蛋白,它以滑动环的方式增加酶的进行性。DNA 聚合酶 δ 合成先导链和后滞链。DNA 聚合酶 α 的功能是在后滞链上延伸 RNA 引物,为**复制因子 C(RF-C)**提供模板,RF-C 的功能可能类似于大肠杆菌 γ 复合物,即 DNA 聚合酶的定位和控制后滞链上复制的进行性。DNA 聚合酶 ε 的功能还不清楚。它的结构类似于 DNA 聚合酶 δ,但不与 PCNA结合。它可能在复制叉上发挥作用,或者像 DNA 聚合酶 β 一样涉及 DNA 的修复。β 是五个酶中最小的酶,其保真度和进行性是最低的。DNA 聚合酶 γ(类似于大肠杆菌 pol I)负责线粒体 DNA 的复制,类似的酶从植物叶绿体中也分离得到。

表 26.6　真核生物 DNA 聚合酶的性质和可能的功能

哺乳动物中的名称	α	β	γ	δ	ε
酵母中的名称	pol1	pol4	PolM	pol3	pol2
酵母基因	*POL1*	*POL4*	*MIP1*	*POL3*	*POL2*
定位	细胞核	细胞核	线粒体	细胞核	细胞核
亚基数目	4	1	2	2	>1
5′→3′多聚酶	√	√	√	√	√
3′→5′外切酶	×	×	√	√	√
引发酶	√	×	×	×	×
相关因子	没有	没有	没有	PCNA	没有
进行性	中等	低	高	有 PCNA 时高	高
功能	后滞链引发	修复聚合酶	细胞器聚合酶	基本复制聚合酶	不清楚

DNA 引发酶　　**DNA 引发酶**(DNA primase)是在 DNA 合成中合成 RNA 引物(参阅)的酶(它们可以用 NTPs 和 dNTPs 作为底物,经常合成混合引物)。引发酶行使两个功能。首先是它们在起点处起始先导链的合成(这需要进行一次);第二是它们在延伸中帮助冈崎片段的反复起始。对细胞基因组,两种功能使用同样的引发策略(见下面的引物和引发部分)。

在大肠杆菌中,DNA 引发酶是 *dnaG* 基因座编码的。对 *oriC* 型复制子的复制(如染色体本身),引发酶的有效引发活性依赖于 Dna B 解旋酶。引发酶和解旋酶形成功能复合物,称为**引发体** (primesome)。噬菌体复制子有不同的引发酶要求。噬菌体 G4 的复制起点被引发酶直接识别,但没有引发体。ϕχ 型复制子(如噬菌体 ϕX174)需要复杂的引发体,它含有 6 个**预引发蛋白**:DnaB(解旋酶),DnaC,DnaT,PriA,PriB 和 PriC。PriA 为起点(或者是引发体装配位点,pas)识别所必需,但 PriB 和 PriC 和 DnaT 的功能还不清楚。这些因子组建形成**预引发体**,被 DnaG 引发酶识别。在两类起点中,引发体是可移动的复合

物,以 ATP 依赖的方式沿复制叉运动方向在 DNA 上移动(也就是说与后滞链合成的方向相反),设置每个冈崎片段的 RNA 引物。

在真核生物中,两个与 DNA 聚合酶 α 结合的多肽有引发酶活性。这是真核和细菌引发体的基本差别,也就是说大肠杆菌的单个基本的复制 DNA 聚合酶(pol Ⅲ)延伸两条链,而在真核生物中不同的酶在后滞链上延伸引物,只有它需要与引发酶结合。细胞引发酶和解旋酶可相互作用,但它们是不同的酶。然而,几种真核病毒编码的蛋白同时有解旋酶和引发酶活性。

DNA 解旋酶　**DNA 解旋酶**(DNA helicases)是沿单链 DNA 移动的酶,它利用 ATP 水解得到的能量打断氢键,分离双链分子。解旋酶是很多细胞过程所要求(参见如*核苷酸切除修复、同源重组、转录终止、接合、Ti 质粒*),也是复制过程中提供单链模板所必需的——它们是进入复制体的第一个成分。解旋酶经常只沿 DNA 一个方向移动,可根据它们 5′→3′或 3′→5′的极性分类。在拓扑学上扭转 DNA 中、解旋酶和拓扑异构酶协同作用释放扭曲的链。有些细菌和真核复制解旋酶的性质在表 26.7 中列出。

表 26.7　复制解旋酶和它们的功能

解旋酶	极性和功能
细菌复制解旋酶	
DnaB 蛋白	5′→3′极性。大肠杆菌中染色体复制主要的解旋酶;dnaB 有快停表型(表 26.2),并且是体外系统中唯一需要的解旋酶
PriA 蛋白(n′,Y)	3′→5′极性。引发体的成分
Rep	3′→5′极性。滚环复制必需。在细胞复制中的作用不清楚
真核复制解旋酶[a]	
DNA 解旋酶 A	从酵母和小牛胸腺中提取。与 DNA 聚合酶 α:引发酶一起纯化
DNA 解旋酶 δ	5′→3′极性。与 DNA 聚合酶 δ 共纯化
DNA 解旋酶 E	3′→5′极性。与 DNA 聚合酶 ε 共纯化
RF-A 相关解旋酶	从不同物种中得到的几个酶,或者是与 RF-A 共纯化,或者被其促进
RF-C 相关解旋酶	与复制因子 C 共纯化

在细菌中,复制和其他解旋酶的功能研究得较清楚。在真核生物中,发现有多于 30 种解旋酶活性,并且在酵母基因组序列中还有另外的假定的解旋酶,但只有少数功能被精确描述。

a 与复制蛋白共纯化的解旋酶。

DNA 拓扑异构酶　DNA 拓扑异构酶是催化 DNA 拓扑异构体(DNA 不同的拓扑异构形式,见第 16 章核酸结构:文框 16.2)相互转换的酶,在复制中需要它们释放解旋酶解链产生的扭曲的张力;它们在转录和重组中行使类似的功能。它们也去除和解离复制和重组中产生的节和环(中间阻断环)。在大肠杆菌中,很多涉及核酸(包括复制、转录、同源重组和可移动元件的转座)的反应是倾向于负*超螺旋的*(参阅)。在真核生物中拓扑异构酶是细胞核骨架的一部分,有利于将 DNA 组织成功能结构域(见第 3 章染色质);它们在有丝分裂中姐妹染色体的分离中也起主要作用。

拓扑异构酶可以分成两类功能类型(表 26.8),其差别主要在于反应的机理:Ⅰ型拓扑异构酶只剪切一条链,因此只能连接/去连接含有缺口的底物;Ⅱ型拓扑异构酶可以同时剪切两条链 ,因此可以连接/去连接共价闭环链 。大多数拓扑异构酶催化超螺旋的释

表 26.8　大肠杆菌和真核拓扑异构酶

	Ⅰ型拓扑异构酶	Ⅱ型拓扑异构酶
剪切	一条链	二条链
ΔL	1 步	2 步
机制	酶非共价与 DNA 结合。一条链被剪切，3′磷酸基团共价与活性位点的酪氨酸残基结合。完整的链从缺口处通过。打断的链重新连接。	酶非共价与 DNA 结合。两条链都被剪切，5′磷酸基团共价与活性位点的酪氨酸残基结合（这涉及对回旋酶和真核拓扑异构酶Ⅱ有 4bp 交错）。一完整的双链从缺口通过（通过包括 ATP 结合的构象改变而得以进行）。打断的双链重新连接，DNA 释放。
需要 ATP	不	是
大肠杆菌	topo Ⅰ（ω 蛋白）	回旋酶（topo Ⅱ）。诱导负超螺旋
	topo Ⅲ	topo Ⅳ。去连接连锁的环
	释放负超螺旋	
真核	拓扑异构酶Ⅰ。释放正和负超螺旋 拓扑异构酶Ⅲ。释放负超螺旋（弱活性），不去连接	拓扑异构酶Ⅱ。释放正和负超螺旋 拓扑异构酶Ⅳ。释放负超螺旋，可能也能释放正超螺旋

ΔL，连接数的改变（参阅）

放，但不能在松弛的 DNA 中引入超螺旋。大肠杆菌 **DNA 回旋酶**（DNA gyrase）是个例外，它可以产生负的超螺旋，从硫化裂片菌属（*Sulfolobus acidocaldarius*）中分离的**反回旋酶**（reverse gyrase）可以产生正的超螺旋。回旋酶的作用似乎是维持 DNA 处于能量有利的状态。反回旋酶的功能更不清楚，它可能与在*S. acidocaldarius*（参见*耐温多聚酶*）生活的高温环境中维持染色体双链构象有关。

单链结合蛋白　　**单链结合蛋白（SSBs）**是缺乏酶活性的复制辅助蛋白，但是复制体其他酶有效活性所必需的（表 26.9）。SSBs 在细胞内行使很多涉及单链区域稳定性的功能（如*同源重组*，也见 *hnRNP 蛋白*）。在复制中，这包括稳定熔解起点，维持螺旋酶的活性，从 DNA 模板上去除二级结构（参见*框架移动保真度*）和抑制核酸酶活性。SSBs 对单链 DNA 有很高的亲和性，但对双链 DNA 和 RNA 没有亲和力。它们与 DNA 协同作用，将 DNA 包被上蛋白聚合体。它有某些碱基组成的倾向性，但很少有顺序专一性结合。

表 26.9　大肠杆菌和真核生物单链 DNA 结合蛋白的性质

	大肠杆菌	真核生物
蛋白	SSB（单链结合蛋白）	RP-A（复制蛋白 A）或 HSSB（人 SSB）
结构	19kDa 亚基的同源四聚体	70kDa，30kDa 和 15kDa 亚基的异源三聚体
功能	复制、重组和修复中稳定单链区域 M13 相关基因组起点直接引发 在引发体复合物中与 PriB 相结合（可能有引发功能）	在复制、重组和修复中稳定单链区域 与 pol α 相互作用：引发酶阻断非特异性引发事件 与转录因子、修复蛋白 XP-A 和几种解旋酶相互作用（在转录和修复中可能有特定的作用）
性质	ssDNA 专一性，但没有顺序特异性。协同结合	ssDNA 专一性，部分顺序专一性。活性通过磷酸化调节
基因	*ssb*	*RP-A1*，*RP-A2*，*RP-A3*

SSBs也直接与复制体的不同成分相互作用以刺激它们的活性。在大肠杆菌和真核生物中,SSB与引发酶或引发体的成分相互作用将促进特定的引发活性。大肠杆菌SSB也可通过覆盖除复制起点外的单链基因组所有位置,指导其引发,复制起点位置是以发夹的二级结构为特征的。噬菌体T4和T7编码的SSBs直接刺激噬菌体编码DNA聚合酶的活性。

核酸酶　核酸酶(nucleases)是通过水解磷酸二酯键消化核酸的酶。核酸酶如果底物是DNA就是**脱氧核糖核酸酶(DNase)**,如果底物是RNA就是**核糖核酸酶(RNase)**。**外切核酸酶**需要游离的末端,逐步消化分子,而内切核酸酶能够水解内部磷酸二酯键,可以将共价闭环的模板作为底物。**切除核酸酶**帮助释放寡核苷酸片段。核酸酶在它们的底物专一性上变化很大。外切酶可以是单链或双链专一性的,可以有特定的极性,可以需要特定末端结构起始消化。内切酶也可以有单链或双链底物倾向性,可以有不同程度的顺序专一性;**切口酶**(nickase)是剪切双链中一条链的内切酶。核酸酶在很多系统中起重要作用(参见限制性内切酶、UvrABC核酸酶、RecBCD核酸酶、反转录病毒。接合、cDNA合成、重组DNA、RNA加工、转座酶、整合酶、核酶、剪接体)。

在复制中需要几种来源的核酸酶活性。一般情况下,通过DNases和RNases回收的核苷酸为复制和转录提供了底物(**补救途径**,salvage pathways),控制这些途径的基因突变产生快停的表型(参阅)。在复制中需要三个特定的核酸酶活性。校正需要3′→5′外切酶活性,这是DNA聚合酶的内在活性(见上文)。引物的去除需要RNaseH(从DNA:RNA杂交链上特异性消化RNA链的酶)和一个5′→3′外切酶活性的组合。在大肠杆菌中,后者是基本修复酶DNA聚合酶I的内在活性,而在真核生物中,它是由外切酶MF-1提供。RNaseH在大肠杆菌中不足以去除引物,在真核生物中推测也是如此,因为它不能剪切连接核糖核酸和脱氧核糖核酸的磷酯键。RNaseH消化引物因此至少留下一个核糖核苷酸残基附着在冈崎片段上,并且由于DNA引发酶对脱氧核糖核酸和核糖核酸的混用,因此可能有几个内部的核糖核酸残基需要被去除。RNaseH也加工ColE质粒复制的先导链引物(见第20章质粒)。

DNA连接酶　DNA连接酶是在双链DNA中催化相邻核苷酸形成磷酸二酯键的酶。它要求5′的核苷酸必须有完整的磷酸基团而3′核苷酸必须有完整的羟基。在细菌中,DNA连接酶需要NAD辅助因子,而在真核生物和古细菌中,连接酶需要ATP。无论是何种情况,辅助因子提供了与酶活性位点共价结合的腺苷基团。然后这个基团转移到5′核苷酸上,接着被邻近核苷酸的3′羟基进攻形成磷酸二酯键。DNA连接酶控制所有DNA修复途径的最后阶段,即双链DNA中一条链的缺口修复。哺乳动物似乎至少有四种连接酶活性,DNA连接酶I是涉及后滞链修复的基本酶(参见噬菌体T4 DNA连接酶)。

细菌和真核细胞复制体的比较　细菌和真核复制体在结构和功能上存在着重要的平行关系:它们不对称的结构、存在高进行性的聚合酶和分步进行的聚合酶分别合成先导链和后滞链,以及类似的酶学性质。它们也存在重要的区别,如后滞链引发和修复的策略。这些策略的比较见表26.10。细菌和真核生物在细胞分裂周期中DNA复制的组织也存在着显著的差异——讨论见细胞周期。

表 26.10 细菌和真核复制体的成分和性质

复制体性质/成分	细菌	真核生物
一般性质的比较		
复制起点	单个	很多
延伸速度	100kbp min^{-1}	2 kbp min^{-1}
冈崎片段	1～2kbp	100～200bp
引发策略	引发酶产生引物，由 pol Ⅲ延伸	引物酶产生引物，由 pol α 延伸，由 pol δ 完成
复制多聚酶	有不同进行性单体的异源二聚体	每条链有不同的酶，有不同的进行性
拓扑学	松弛和卷曲之间存在平衡，净负超螺旋	浓缩在核小体中的负超螺旋。高级结构影响拓扑学
复制体成分的比较		
复制多聚酶	pol Ⅲ全酶	pol δ/pol α
进行性因子	β 亚基	PCNA
定位因子	γ 复合物	RF-C
引发酶	DnaG	polα：引发酶
解旋酶	DnaB(定位需要 DnaC)	? 几个候选者
引物去除	RNaseH 和 pol Ⅰ	RNaseH1 和 MF-1 核酸酶
后滞链修复	pol Ⅰ和 DNA 连接酶	polδ/polε 和 DNA 连接酶Ⅰ
拓扑异构酶	DNA 回旋酶	topo Ⅱ
单链结合	SSB	RF-A

26.3 复制的起始

起始的一般原则 基因组复制开始于一个或多个顺式作用位点，称为**复制起始点**(简单地称为起点)。它们通过提供复制所需蛋白因子和酶的装配位置而增加起始的效率，否则这些因子和酶将随机与 DNA 结合。在起始的静态蛋白集合称为**起始体**(orisome)，当它开始从起点移动时成为复制体。起点也提供复制调节的靶位点，因此起始是复制调节的主要阶段。复制起始被起始因子所控制，但也可以受其他蛋白和 DNA 本身的性质影响。起点可以通过用电镜检查标记的复制中间体，通过标记和分离标记新生 DNA 链或者通过双向电泳分离 DNA 结构来鉴定其物理位置(文框 26.2)。与很多其他顺式作用元件一样，起点可以通过体外诱变(参阅)功能性作图。

在细菌染色体、病毒基因组和质粒中，单个起点被用来起始复制[①]。这样的元件构成了单个**复制子**(replicon)，定义为在保守序列控制下的复制单位。更大的真核生物染色体有多个起点，可以被看成是复制子的串联排列。

起点的性质和起始的策略 **复制模型**(表 26.11)依赖于起点的性质和分布及其起始策略。在双链 DNA 基因组中，起点是双链开始解链的位置 。在大多数情况下，随着引物合

①很多质粒和病毒可以有几个选择性的起点，但只有一个在复制循环使用。连续复制的双链 DNA 基因组(如线粒体 DNA)有两个起点，每条链上一个，可以看成两个复制子。

表 26.11 复制模型,反映了起点定位和解旋酶定位的性质

复制模型	机制	举例
半保留复制机制		
偶联双向的	解旋酶定位到两条链,组建两个复制复合物	*oriC*
序列双向的	一解旋酶定位于起始位置,促进单向复制直到第二个引发位置暴露出来,开始相反方向的复制	R6K 质粒
单向的	起点只有一个解旋酶定位,不连续合成链,在起点处终止	ColE1 质粒
取代机制		
单向取代	在解旋酶定位前起点熔解,并被内切酶剪切	pT181

成而发生解链,尽管对于滚环复制,一条链被切开提供了引物末端。对单链 DNA 基因组,解链和切口并不是必需的,起点只是引物合成的位置。在每种情况中,起点的精确位置对应于新生 DNA 合成的开始,是引物末端产生的位置。在实际上,起点经常不能准确定位:引物长度在一个短的范围内变化(8～10 个核苷酸),因为 DNA 引发酶对核苷酸使用的不专一性,引物经常由混合的核糖核酸和脱氧核糖核酸亚单位组成。在 RNA 复制中,起点与启动子同义(参阅),也就是说 RNA 复制酶能够与其底物结合并起始合成的位置。

对双链 DNA 基因组的半保留复制(所有细胞基因组和很多质粒和病毒)而言,在起点处双链打开建立一个或二个复制叉。解链经常被特定的起始蛋白所调节,它们征集解旋酶和引发酶,尽管在有些体系中(如 ColE1 相关的质粒)RNA 聚合酶行使这一功能并有引物合成的作用。解旋酶对亲本链的连续解链是必需的,起始蛋白的一个重要功能是将解旋酶征集到复制叉中(偶尔起始蛋白本身有解旋酶活性,如 SV40 T 抗原)。在大多数起点处,解旋酶定位到两条链上,产生两个复制叉,复制是双向的。在有些质粒体系中,单个解旋酶定位且复制叉是单向的。经常先导链被引发,延伸一段距离后后滞链的第一个冈崎片段在反向模板上被引发。

对于连续复制只需要单个的引发事件(对双链基因组每条链上一个)。单链 DNA 基因组的引发位置经常通过其二级结构识别,对于线性的质粒和病毒基因组而言,起点是每条链的 3′末端。在双链基因组中,建立取代复制叉允许在一个方向连续合成,不断取代一亲本链。在有些阶段,被取代链上另一个起点暴露出来,使第二链的合成可以开始。这个策略在线粒体 DNA 复制和腺病毒和有末端蛋白的几种相关线性质粒的复制中使用。

细菌复制起点　大肠杆菌染色体起点①定义为 *oriC*,它对正常复制非常重要。最小的序列是 245bp 长,含有一系列串联重复元件,即含有四个九聚体和三个 13 聚体的**重复区**(iterons)。九聚体是 DnaA 起始子的结合位点,分布在"识别位点"上;三个 13 聚体往往在富含 AT 的区域邻近。*dnaA* 突变等位基因是慢停表型,表明 DnaA 蛋白的功能局限于起始。DnaA 以 ATP 依赖的方式与它的识别位点协同结合,所以最后出现 30 个 DnaA 分子。DnaA 的结合促进在低扭曲稳定性的三个 13 聚体处双链解旋。然后 DnaBC 解旋酶复合物定位,这个过程被作用于两个邻近启动子上的 RNA 聚合酶作用所促进(参见细菌人工染色体)。起始序列在表 26.12 中总结。

①细菌质粒和噬菌体复制起点在其他部分讨论——见质粒、病毒各章。

表 26.12 大肠杆菌复制起始的阶段

起始阶段	性质
起始复合物	DnaA 与九聚体 DnaA 盒结合 不需要 ATP
开放复合物	DNA 在 13 聚体处熔解 需要 ATP 和控制 DNA 结构的 HU 蛋白(参见类核)
预引发复合物	DnaBC 解旋酶定位。DnaC 从复合物中释放
引发复合物	DNA 被 DnaB 解链,加上引发酶和 pol III
延伸	征集拓扑异构酶和 SSB,延伸开始

oriC 一个显著的特点是存在 14 个 Dam 甲基化位点。半甲基化的起点是复制起始后直接产生,不能促进重新起始,并与细胞膜结合。有可能 DNA 的甲基化(参阅)有助于在细胞分裂中控制复制的时间。上面描述的排列结构在其他革兰氏阴性细菌的复制起点中中高度保守。重复区的序列以及它们的空间排列看来都很重要的。在革兰氏阳性细菌中,序列有差异,但结构保持相同。

滚环起点和 θ 起点的结构不同,尽管它们都有起始蛋白的结合位点和易熔解的 AT 丰富元件。熔解暴露了链剪切位点,从而产生了引物末端。除了这些区别,在所有细菌复制子中起始的过程是类似的,包括起点的识别和解旋酶定位后的解链。

真核系统的起点 简单真核体系(质粒、病毒和单细胞真核生物)的起点类似于原核起点的模式(表 26.13)。通常,它们含有一个核心序列(***ori*-core**),*ori*-core 含有起点识别元件,一个 DNA 解链元件和 AT 丰富元件。在核心序列外有一个或多个辅助元件,是转录因子的结合位点。它们被认为是以影响转录起始同样的方式增加了复制起始的效率,即通过与系统核心成分,**起始体**(orisome)的相互作用(参见基本转录装置)。简单的起点可作为**自动复制序列**(autonmomously replicating sequences,**ARSs**)起作用,也就是说如果在细胞中一个序列与任何其他 DNA 片段连锁时将赋予它自动复制的能力,那么连锁的 DNA 片段从该序列中获得了 ARS(参见酵母克隆载体)。

表 26.13 简单真核生物起点的成分

起点成分	功能
起点核心	
起点识别元件(ORE)	起始因子开始结合的位点——有助于 DNA 解链和复制体成分的定位
DNA 解链元件(DUE)	DNA 解链和复制体进入双链的位置
AT 丰富元件	与简单起点相联系的基序,有 T 丰富和 A 丰富的链。容易发生弯曲,有助于在 DUE 处 DNA 的熔解和 ORE 与起始蛋白的相互作用
起点辅助	
辅助元件	转录因子结合位点,如果与转录因子结合则增加复制起始效率。可能用转录因子控制复制的调节,使复制能以细胞类型专一性的方式和发育程序化方式调节(不同的激活可以控制转录活性和抑制染色质)早期和晚期复制。转录因子能够通过帮助起始蛋白的结合、改变它们的活性或调节染色质的结构或 DNA 的结构来增强复制。在线粒体 DNA 中,它们可以在起点产生引物支持转录

多细胞生物染色体起点 与简单起点比较,后生动物染色体起点分析较困难。功能研究表明 DNA 的长度对复制是关键因素,而不是任何特异的序列。然而通过 PCR 鉴定复制起始处的序列经物理分析表明这涉及序列特异性。DNA 开始合成的位置,**初始起点**

(或者**双向复制起点**——大多数初始起点是双向的)两侧有一些**二级起点**。初始起点经常为 0.5～2kbp 长,而二级起点可跨越 50kbp,因此称为**起始地带**(initiation zone)。

已知的后生动物复制起点,只有一些当克隆到质粒中时方证明有 ARS 活性,因此染色体结构被认为在起点功能中有重要的作用。在起点两侧经常发现转录因子结合位点,尽管转录本身对起始不是关键的。然而,这些因子为起始因子组建提供开放的染色质结构域,也可能直接帮助组建的过程——这直接类似与它们在转录中的功能。尽管简单起点的 AT 丰富元件没有报道,潜在的 DNA 解链元件在几个起点处都被发现。还存在具有形成功能还不清楚的十字形结构的潜在能力的二重对称位点。也有细胞核基质附着位点(参阅)。

两个棰头结构起点的显著特点是 100～500bp 的**高密度的甲基化岛**(densely methylated island, **DMI**),其中两条链所有的脱氧胞嘧啶都被甲基化。这不是哺乳动物中经典的 CpG 甲基化模式,因为甲基化胞嘧啶的前后序列似乎不是很重要(参见 CpG 岛)。这些起点甲基化功能还不清楚,但它可能影响染色质结构或转录因子的结合,或者它可能调节复制的时间,如同在细菌中的情况。

26.4 引物和引发

引发的基本原理 DNA 合成对引发的需要可能反映了校正功能的进化。校正是通过 DNA 聚合酶不能延伸不与模板配对的引物来实现的,但在每次链合成的起始中,因为没有引物来延伸,酶为了开始合成,必须容忍一些不精确性。对于先导链的合成,突变会在起点处积累。然而,对于后滞链的合成,一个没有引物的机制将在整个基因组中导入周期性的突变。引发以一小段 RNA 开始,使复制体能够确定起始可以发生的不精确区域,并当有 DNA 引物末端存在时,以 DNA 来替代它们。这样引发保持了基因组的完整性。

引发策略 在细胞 DNA 引发酶合成的 RNA 引物中被用于先导链和后滞链的合成。通常引物或**起始 RNA**(initiator RNA, iRNA)长约 10 核苷酸。在原核生物中,引物直接由 DNA 聚合酶Ⅲ延伸,而在真核生物中,DNA 聚合酶 α 在每个引物加上 4～5 脱氧核苷酸[这称为**起始 DNA**(initiator DNA, iDNA)],DNA 引物末端由 DNA 聚合酶 σ 延伸。DNA 引发酶合成的短 RNA 引物是后滞链合成的通用引发机制。非细胞复制子发展了一系列不同的先导链起始引发的策略(表 26.14)。

表 26.14 在复制起点上先导链引发的机制

引发机制	细节	举例
RNA 引物	DNA 引发酶合成的短新生寡核苷酸 RNA 聚合酶合成的特殊转录本 预先形成的 tRNA 引物的退火	细胞 DNA 线粒体 DNA,ColE 质粒 逆转录病毒的反转录
发夹引发	有反转的末端重复顺序的单链模板可以反转形成发夹	疱疹病毒
内切引发	内切酶剪切双链 DNA 产生缺口。暴露的 3′-OH 可以作为引物末端	很多革兰氏阳性细菌质粒的滚环复制
入侵引发	一 3′末端通过同源重组导入完整的双链	T4 晚期复制
末端蛋白引发	有些线性复制子由末端核酸结合蛋白帮助引发	腺病毒
寡核苷酸引发	在体外延伸时寡核苷酸退火到模板上	聚合酶链式反应、随机引发、cDNA合成(参阅)

26.5 复制的终止

环状模板复制的完成　　对于大肠杆菌染色体和有些质粒，环状基因组的复制可以在**终止区域**(terminus region，*ter*)的特定序列终止。此外，当两个复制叉从相反方向前进相遇时，复制可以被简单地停止，如SV40基因组。大肠杆菌染色体的终止区域含有两对反向重复顺序。它们与抑制复制叉前进的结合蛋白Tus结合，解旋酶活性的抑制促进了这一过程。终止区域将大肠杆菌基因组分成两半，称为**复制分布**(replichores)，每个都在双向起点开始而在末端终止。*ter*-Tus阻断复制的能力是方向依赖性的。特定终止策略的优点还不清楚：在大肠杆菌中缺失 *ter* 位点或 *tus* 基因都没有影响。终止可能提供了互锁环去连接的调节机会。

线性模板复制的完成　　线性基因组需要特殊的策略来完成它的5′末端。DNA聚合酶只能以5′→3′方向合成DNA，并需要一个预先形成的引物，因而不能完成每个后滞链尽头的5′末端，因为它没有上游供引物结合的位置(图26.4)。没有特殊的完成策略，染色体在每轮复制后将变短。环形基因组在这方面有优势，很多线性复制子通过在它们的复制循环中采取一环形构象来解决末端问题(如λ噬菌体)。进化产生了很多涉及后滞链的合成更精细的策略，被列在表26.15中。

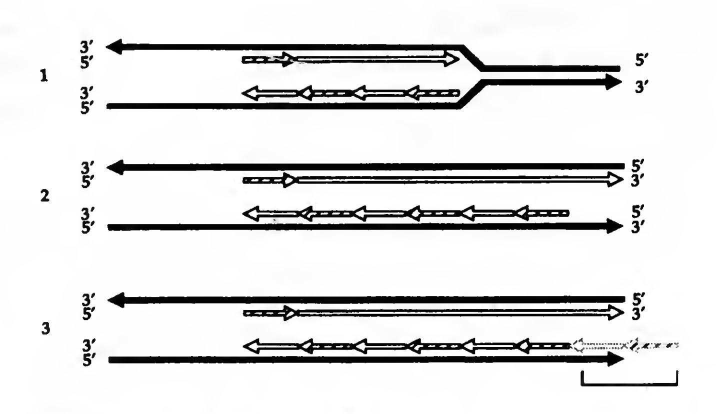


图26.4　线性模板末端的复制问题。(1)向右侧进行的复制叉有一条先导链(上)和一条后滞链(下)。亲本链是黑色，新生链是白色，引物是阴影的；注意在先导链上只有一个引物而在后滞链上有多个引物。(2)先导链的3′末端到达模板的5′末端。其间最后的后滞链引物被合成，后滞链上最后的冈崎片段被引发。(3)没有位置供下一个引物结合，所以后滞链的最后部分没有完成，染色体保持在(2)的状态。

真核生物通过在染色体末端加上核苷酸形成端粒(参阅)来解决复制问题。端粒的重复顺序是通过特殊的酶，端粒酶(参阅)添加来维持染色体长度。最近产生的端粒酶基因敲除小鼠表型是正常的，但突变的效应在六代后出现，存在的端粒渐至完全缺失的。除了端粒基因的缺失效应，突变的小鼠表现出染色体突变，如融合和异位，反映了端粒的另一个功能——区别自然染色体末端和随机的断裂。对端粒的进一步讨论，见染色体结构和功能。

表 26.15 完成线性基因组 5′末端的策略

策略	机制和举例
共连接	线性基因组有冗余末端,允许进行环化或共连接。这样的结构可以被剪切产生有 5′突出末端的单个基因组,这样的末端可以被常规 DNA 合成补平。如噬菌体 T7 和 λ[1]
末端蛋白引发	以末端蛋白起始链合成的病毒不需要特殊的末端机制:它们在 5′末端起始每一条链,能够在 3′末端完成,如腺病毒,噬菌体 φ29
发夹引发	从链额外 5′末端起始链合成的引发策略
共价补平策略	有些病毒表面上是线性双链,有一个共价补平的末端,因此熔解产生一个单链环,可以象环形复制子一样复制,如类病毒
端粒	称为端粒酶的酶将寡核苷酸加到线性染色体的末端。尽管在这种方法中额外的 5′顺序丢失,复制后的端粒酶活性可以补充端粒,所以在连续复制循环后没有真正的基因丢失(见第 5 章染色体结构和功能)

26.6 复制的调节

复制的时间控制 所有的细胞将基因组复制与细胞周期相协调,以阻止 DNA 的丢失或过剩(见第 2 章细胞周期),而质粒的复制经常是拷贝数依赖的(见第 3 章质粒)。延伸速度慢和基因组大的真核生物,通过增加每条染色体上有功能的起点来加快复制(如爪蟾发育中的剪切阶段)。真核生物复制是时间控制的,所以不是所有的起点同时激活,基因组不同区域以被调节的时间顺序复制。复制的时序调节与遗传活性相对应,也就是说管家基因复制较早,而细胞类型特异的基因在它们表达处一般复制早,而在转录沉默的细胞中复制较晚。时间控制是发育调节的:很多细胞类型特异的基因在胚胎中是晚复制的,但是当合适的细胞类型分化后转变成早复制;X 连锁基因当 X 染色体失活(参见 *X 失活*)后从早复制转变成晚复制。

复制时间的区域性质可以通过 S 期中脉冲标记染色体来研究。早期复制 DNA 用溴脱氧尿嘧啶标记,产生与淡 G 带和 DNase Ⅰ敏感区域相一致的带型[参见*染色体带型*、*等分布(isochore)*]。因此早期复制的 DNA 似乎是富含基因染色体的活性区域。对前中期染色体高分辨率带型分析显示了将主带分成 1～2Mb 长度区域的亚带带型。哺乳动物染色体复制子长度为 50～300kb,说明每个复制时间区含有许多同步的起点,这被通过放射自显影观察到的串联起点所证实。

易位表明复制的时间不是很多特定起点的内在性质,而是位置依赖性的。因此时间的控制可能由顺式作用位点调节。几个候选元件已经被发现,包括端粒序列和 β 珠蛋白基因座的*基因座控制区域*(参阅)。在哺乳动物中,可能因为与细胞核基质的普遍相互作用,时间相同的复制叉在细胞核中作为**复制焦点**(replication foci)彼此成簇。复制时间、染色质结构、DNA 甲基化和转录活性的相互关系已经建立。然而这些成分的因果关系还不清楚。

---

1)λ 噬菌体是以 θ 型和 σ 型两种形式进行复制的,不存在线性 DNA 的复制,故此例可能有误。——译者注

文框 26.1 核酸合成

模板指导的核酸合成　在**模板指导的合成**(template-directed synthesis)中,原先存在的核酸作为模板将核苷酸组装形成新的链,通过碱基配对指导掺入的序列。这是核酸信息转移的基础。有两种类型的核酸,DNA 和 RNA,那么就有四种类型的核酸合成反应,每个由特定的**模板依赖核酸聚合酶**(或合成酶)催化,如下文所述。这些酶在模板使用方面是特异,酶本身决定了底物的专一性。

如表中所使用的,DNA 聚合酶和 RNA 聚合酶的名称是令人误解的。一个 **DNA 聚合酶**(DNA polymerase)是限制在 DNA 合成活性的酶(也就是说这个名称包括了反转录酶和末端脱氧核苷酸转移酶和 DNA 复制中的 DNA 聚合酶)。同样 **RNA 聚合酶**(RNA polymerase)是有 RNA 合成活性的酶(包括转录酶、病毒 RNA 复制酶,DNA 引发酶和多腺嘌呤聚合酶)。DNA 聚合酶也称为 **DNA 复制酶**(DNA replicase),而 RNA 聚合酶也称为**转录酶**(transcriptase)。一种酶可以有多种聚合酶活性,如大多数转录酶有另外的 DNA 依赖的 DNA 聚合酶活性。

聚合酶反应　所有模板指导的核酸合成反应涉及称为**模板引物**(template primer)的结构。引物是新生链,之所以这样呼法是因为它通过提供延伸的末端引发连续延伸;模板链提供了引物延伸的信息。反应的底物是核苷酸三磷酸:延伸涉及引物链末端 3′羟基对掺入核苷酸的 α 磷酸基团的亲核攻击。这样形成一 5′→3′磷酯键,释放焦磷酸,表示成:

$$NA_x + NTP \longrightarrow NA_{x+1} + PPi$$

拷贝过程	酶:系统定名(普通名称)
DNA ⟶ DNA(**DNA 复制**)	DNA 依赖的 DNA 聚合酶(**DNA 聚合酶**)
RNA ⟶ RNA(**RNA 复制**)	RNA 依赖的 RNA 聚合酶(**RNA 复制酶**)
DNA ⟶ RNA(**转录**)	DNA 依赖的 RNA 聚合酶(**RNA 聚合酶**)
RNA ⟶ DNA(**反转录**)	RNA 依赖的 DNA 聚合酶(**反转录酶**)

在体外系统中加入过量的焦磷酸使反应向反向进行(**焦磷酸解作用**,pyrophosphorylysis)。在体内焦磷酸通过转变成无机磷酸从细胞中除去,使反应不可逆。通过形成 5′→3′磷酸二酯键的延伸使链只能以 **5′→3′方向**生长,这称为**尾的生长**(tail growth)。聚合酶反应可以是**进行性的**(如果不释放酶进行很多多聚化步骤)或者是**分散性的**(如果酶在每次掺入后解离)。聚合酶经常被发现与其他蛋白相结合,通过调节酶或模板结构来增强酶的进行性。

DNA 和 RNA 合成的差别　尽管所有的核酸合成反应都是类似的机制,DNA 合成与 RNA 合成有三个方面的重要差异。首先使用不同的底物——DNA 合成使用脱氧核糖核苷三磷酸而 RNA 合成用核糖核苷三磷酸。第二,在起始阶段,RNA 聚合酶可以是从头开始反应,也就是说酶可以在没有预先存在引物的情况下在模板对应位置掺入第一个核苷酸。。RNA 链的第一个核苷酸保留它的 5′三磷酸,作为进一步延伸的引物。相反 DNA 聚合酶反应不能从头开始,需要一个已经存在的引物。所有的 DNA 聚合酶(包括反转录酶)都需要引物,只有一个例外是脉孢菌属中的线粒体逆转录酶。在细胞中,引物是在复制中心由指定的酶合成的短 RNA 分子。对其他基因组有不同的引发策略(表 26.14)。引物增加了 DNA 复制的保真度,但使线性产物 5′末端的完成复杂化(见正文中的讨论)。最后,因为 DNA 合成是基因组复制的基本方式,而 RNA 合成是基因表达的基本方式,很多(不是所有)DNA 聚合酶有相关的内切酶活性进行校正——在引物末端鉴定和去除错配核苷酸。RNA 聚合酶没有这样的活性。

非模板核酸合成 在不依赖模板的合成(template independent synthesis)中,核苷酸不需要模板的指导,就可以添加到已经存在的核酸末端。这样的反应分成两类:底物由酶选择的专一性反应和可使用任何核苷酸产生随机序列的非专一性反应。一些不依赖模板的酶和它们的细胞功能在下表中列出。

末端脱氧核苷酸转移酶(TdT)	负责在V-D-J重组中在接头处加上额外核苷酸(见第25章重组);在DNA3′羟基上非特异性加上脱氧核苷酸(参见**标记**、**同源多聚尾**)
多腺嘌呤聚合酶(PAP)	负责在真核生物mRNA 3′末端加上腺嘌呤残基(参见**多聚腺苷酸化**)
mRNA鸟苷转移酶	负责真核生物mRNA 5′末端加上7-甲基鸟嘌呤帽(参见**加帽**)
多核苷酸磷酸酶	细菌非特异核苷酸多聚酶。特点是用核苷二磷酸而不是三磷酸作为底物

文框 26.2 复制策略和复制中间体

复制中间体 复制中间体是当基因组部分复制时观察到的结构(参见**复制中间体**)。它们可以通过将放射性核苷酸掺入到复制的DNA中,然后分离复制子并通过放射自显影观察到。复制中间体的分析能够确定复制的机制,也能用于对复制起点作图。

θ和σ结构 双链DNA的环状复制子可以形成四种主要类型的中间体,分属θ结构和σ结构。**θ类型结构**是在内部起点起始产生的。经典的θ结构与希腊字母θ形状相似,表明半保留复制,半保留复制可以称为**θ复制**或**Cairns复制**。类似的结构可以通过内部取代复制产生,尽管被取代的环没有被标记,所有中间体在第一轮复制中表现为一个环,在第二轮中表现为θ型。这称为**取代环(D环)**,类似有自由DNA链的重组中所看到的(见重组)。**σ类型结构**也有两种变化。经典的**σ结构**与希腊字母σ形状相似,表明滚环复制,字母的尾巴代表了伸出的连接链。滚环复制有时称为**σ复制**。在有些进行滚环复制的复制子中,取代链通过一个蛋白保持与新生链的联系,形成双环的中间体,称为**索套结构**(lariat structure)。

线性复制元的中间体 内部起始的早期复制根据是环状复制子还是线性复制子产生一**复制眼**(replication eye)或**复制泡**(replication bubble)(参见**转录泡**)。然而在线性复制子中,内部泡可以相遇产生更大的泡,或者到达末端产生**Y结构**。一个Y结构也可以通过当一条链剥出后在内部取代产生,尽管这与D环一样,在第一轮复制中不能观察到,因为取代链没有被标记。

(陈 淳 译)

第 27 章
RNA 加工

基本概念和定义

- **RNA 加工**(RNA processing)描述了新合成的 RNA 分子在结构和化学方面的成熟。这些修饰过程发生在转录过程中,[**共转录修饰**(cotranscriptional modification)]和转录后[**转录后修饰**(posttranscriptional)];RNA 加工或是对 RNA 的功能极其重要或是代表了一类基因调控的机制。这类反应分成 10 类,见表 27.1。
- 从 DNA 模板复制的 RNA 分子称为**转录本**(transcript)。在转录过程中,产生的是**新生转录本**(nascent transcript),当转录完成后为**初级转录本**(primary transcript)——它是被转录 DNA 的一个精确拷贝。经过一些修饰后,产物为**成熟转录本**(mature transcript),不再与 DNA 完全一样。
- 并不是所有的 RNA 都要经过修饰:细菌 mRNA 很少加工(实际上往往在转录完成之前蛋白合成就开始了),真核 生物的 5S rRNA 转录完成后就有成熟的末端。完全加工和有功能 RNA 分子的前体称为**前体 RNA(pre-RNA)**。真核生物的前体 mRNA 又称为**核内不均一 RNA**(heterogeneous nuclear RNA, hnRNA)。它不同于其他形式的核内 RNA,其分子大小变化很大,反映了基因大小的差异和部分加工的分子的存在。
- 在真核生物核内,转录和 RNA 加工不是在核中自由发生的,而是在核基质中固定的位置进行。特定的 mRNA 加工因子(特别是剪接体和多腺苷酸化酶)与延伸复合物中 RNA 聚合酶Ⅱ的 C 末端作用,所以转录和加工是相关联的。mRNA 加工复合物在核基质中的定位可能有利于产物的输出。在核内和细胞质中 RNA 和蛋白质相结合,形成核糖核蛋白颗粒。

表 27.1 RNA 加工反应的分类

加工反应	举例
剪切	使 rRNA 和 tRNA 从多顺反子转录本中释放 真核生物 mRNA 转录的终止
内切降解	加工 rRNA 和 tRNA 产生成熟末端
核苷酸转移	转移 CCA 到某些 tRNA 的 3′末端
碱基化学修饰	mRNA 和 rRNA 中偶发的甲基化 tRNA 和 snRNA 中的普遍碱基修饰
核苷酸切除和替换	tRNA 鸟嘌呤的高度修饰产生二氢尿嘧啶(Q)和假尿嘧啶(W)
加帽	在真核生物 mRNA 5′末端加上 7-甲基鸟嘌呤
多腺苷酸化	在大多数真核生物和少数细菌 mRNA 3′末端加上多腺苷酸尾巴
剪接(转接)	去除大多数内含子(经常是顺式,偶尔反式)
剪接(连接)	去除 tRNA 内含子
编辑	通过碱基修饰改变 mRNA 所携带信息,在编码区中插入或缺失碱基

27.1 非翻译 RNAs 的成熟

tRNA 剪切和成熟　一些 tRNA 基因可以被单独转录，其他则作为多顺反子的一部分转录。在大肠杆菌中，一些 tRNA 与 rRNA 基因在共同的操纵子中一起转录。根据其来源不同，**前体 tRNA** 可能要经过七种不同的加工反应。多顺反子 tRNA 或 tRNA 和 rRNA 混合转录本中，单个的 tRNAs 通过在成熟的 5′端（经常是鸟嘌呤）切除而释放。在大肠杆菌中，这个功能由核糖核酸酶 P 完成。未成熟的 3′末端由外切酶（可能涉及核糖核酸酶 D）降解，直到出现三核苷酸基序 CCA。如果没有 CCA 的基序，那么就通过 **tRNA 核苷酸转移酶**（tRNA nucleotidyltransferase）在分子的 3′末端加上三核苷酸。所有成熟的 tRNA 末端都有 CCA。在真核生物和古细菌中，有些 tRNA 基因含有需要切除的内含子，这就涉及剪切、末端修饰和 RNA 连接（参见 tRNA *核内含子*）。前体 tRNA 还需要碱基修饰加工过程：tRNA 中的碱基大多数是*主要碱基*（参阅），但大约 10% 在 tRNA 合成中被修饰，通常在原位通过转录后化学修饰完成。鸟嘌呤高度修饰的衍生物产生二氢尿嘧啶和假尿嘧啶，通过类似与 DNA *中切除修复*（参阅）的核酸切除和置换反应而插入。核内小分子核 RNA 中的尿嘧啶也被广泛修饰，所以也称为 **U-RNA**。

rRNA 的剪切和成熟　细菌和真核生物都合成多顺反子 rRNA 转录本。在细菌中有七个 rRNA 操纵子（*rrn*），包含编码所有三种 rRNA 的基因以及某些 tRNA 基因。16S 和 23S **前体 rRNA** 通过碱基配对形成茎环结构。这可能是核酸酶Ⅲ的底物，核酸酶Ⅲ对茎部剪切后释放单个前体。成熟的末端是通过外切酶的加工完成的。

在哺乳动物中，由 RNA 聚合酶Ⅲ转录成单顺反子 5S rRNA 单位，转录成一成熟分子，不需要加工。RNA 聚合酶Ⅰ合成的 45S 的 pre-rRNA，包含 5.8S，18S 和 28S 的 rRNAs。偶尔在多顺反子的转录本中核糖部分被甲基化，其中大多数在成熟 rRNA 中保留，说明它们的功能可能是决定转录本的哪部分保留，哪些部分需要去除。在合成过程中，45S rRNA 在核内与核糖核蛋白复合物缔合，加工也在此处发生。这样的复合物称为**加工体**（processosome）或**核内小分子核糖核酸蛋白**（snorp）（后者是 **small nucleolar ribonucleoprotein**，snoRNP 的缩写，并包含 U3 snRNA）。加工中间体的大小表明外切剪切产生了成熟的终点和必须进一步加工的初步片段。在剪切的同时，一些低等的真核生物 rRNAs 含有自身剪接的内含子，要产生有功能的 rRNA 这些内含子所必须被去除。

27.2 mRNA 的末端修饰和甲基化

细菌和真核生物 mRNA 的加工　细菌的 mRNA 一般不稳定，很少被修饰——它经常在几分钟内完成合成、翻译和降解的过程。

相反，真核生物的 mRNA 较稳定，转运前它在核内被广泛地加工。真核生物的 pre-mRNA 可以进行几种类型的加工反应：末端的加帽和多腺苷酸化；内部的剪接和编辑等修饰；内部腺嘌呤甲基化的化学修饰（这种加工的功能还不清楚）。真核生物的 mRNA 前体在核中并不是以裸露的核酸形式存在，而是与很多的高丰度蛋白缔合形成**不均一的核糖核酸蛋白体**（heterogenous ribonucleoproteins，hnRNPs）。交联有助于我们分析这些蛋

白;它们有 RNA 结合的基序(参见核酸结合蛋白),但以不同的专一性与 RNA 结合,反映了对不同碱基组成的倾向性。蛋白可能以与单链 DNA 结合蛋白(参阅)相同的方式作用,消除 RNA 的二级结构,使其与剪接装置的成分相互作用。它们也可以作为特殊的 RNA 加工因子如剪接蛋白的特殊停泊位点等。

加帽　一旦转录起始,新生真核生物 mRNA 通过在 5′端加上倒转的鸟核苷酸三磷酸而**加帽**(在细菌中还未见加帽的例子)。这个快速反应由**鸟苷酸转移酶**(mRNA guanyltransferase)催化,产生一特殊的 5′→5′磷酸二酯键。这个酶可能与 RNA 聚合酶Ⅱ起始复合物中的成分结合,因为不仅 mRNA 要加帽,RNAP Ⅱ转录的 snRNA 也被加帽(RNA 聚合酶Ⅲ转录的 snRNA 如 U6 snRNA 不加帽)。此结构(5′帽子或简单地称为帽子)是通过**鸟嘌呤甲基转移酶**(guanine methyltransferase)在 G^7 位甲基化,产生 **0 型帽子**(type zero cap),这在酵母中占绝对优势。在更高等的真核生物中,在下一碱基的核糖 $O^{2'}$ 位置加上甲基(转录本中第一个残基,对应位置为 +1),产生 **1 型帽子**(type 1 cap)。如果这个位置是腺嘌呤,N^6 位的碱基也可能被甲基化。在有些物种中,+2 位的碱基也被甲基化,同样也是在核糖的 $O^{2'}$ 位置上产生 **2 型帽子**(type 2 cap)。

真核生物中 5′端的帽子对几种 RNA 的功能是很重要的。它是运出核孔所必需的;对于核糖体结合,防止 RNA 5′端降解也很重要,这可以解释为什么几乎所有真核生物的 mRNA 都是单顺反子(参见蛋白合成,反式剪接,内部核糖体进入位点)。加帽反应可以用来调节蛋白合成,一些动物中卵细胞成熟过程就使用这种策略。大多数 RNA 病毒对它们的基因组和 mRNA 加帽,但是细小核糖核酸病毒的感染策略是避开缺乏加帽的能力,在它们的基因组的 5′末端加上病毒蛋白。正粘病毒也不能在基因组中加帽,但可以从宿主 mRNA 中偷取帽子,这样的过程称为**抢帽**(capsnatching)。

多腺苷酸化　真核生物中 RNA 聚合酶Ⅱ转录终止发生在成熟转录本的 3′端下游几千碱基之后,其精确机制目前还不清楚。3′端通过内切,接着进行**多腺苷酸化**(polyadenylation),加上不同数量(经常为 200 个左右)的腺嘌呤,形成**多腺嘌呤尾巴**(polyadenylate or polyA tail)。在高等真核生物中,多腺苷酸化发生在高度保守的**多腺苷酸化位点**(AAUAAA)后 10～30 个碱基。酵母中的多腺苷酸化位点变化较大。

剪切和多腺苷酸化是由多亚基的复合物进行,复合物中三聚的**剪切多腺苷酸化专一性因子**(cleavage polyadenylation specificity factor, **CPSF**) 识别多腺苷酸化位点,由两个**剪切因子**(cleavage factor) 组成的内切酶,进行剪切反应,**多腺嘌呤聚合酶**(polyadenylate polymerase, PAP)催化加腺嘌呤复合物中还有其他一些未知成分。这些成分中,被认为包含聚合酶Ⅱ的延伸成分,并与其磷酸化的羧端相作用。多腺苷酸化起始时进行较慢,因为 PAP 每加上一个腺嘌呤后都要解离。但当形成一小段多腺嘌呤后,另一成分**多腺嘌呤结合蛋白**(polyadenylate binding protein,PABP)与尾巴结合并加快了 PAP 的进度。PABP 通过一未知的机制,控制多腺嘌呤尾巴的最大长度。

多腺苷酸化具体的功能还不清楚。它可能影响转录本的稳定性,在一些例子中,它在翻译过程中起重要作用。例如果蝇的 bicoid mRNA 直到受精后,三种蛋白促进多腺嘌呤尾巴延伸后才开始翻译。很少真核生物的转录本不进行多腺苷酸化,最值得一提的是组蛋白的 mRNA 和一些植物病毒的基因组。组蛋白的转录本的二级结构负责 3′端的成熟,

这涉及 U7 snRNA 和相关蛋白。大多数细菌没有多腺嘌呤尾巴，但在一些物种中也发现了短的和相对不稳定的寡聚腺嘌呤尾巴，如 *Methanococcus vannielii*。

利用多腺苷酸化，与其配对的合成多聚胸腺嘧啶可在体外扩增或纯化 RNA。用这种纯化方法，得到的大部分是 mRNA，称为**多腺嘌呤 RNA 组份**[poly(A)$^+$ RNA fraction]，其余成分为不含多腺嘌呤成分，如 rRNA，tRNA 等。因为 mRNA 在总 RNA 中比例很小(<5%)，所以这一技术在纯化 mRNA 中很有用(如转录分析，cDNA 克隆，反转录 PCR)。

27.3 RNA 剪接

内含子和剪接反应 内含子是高等真核生物基因中主要的非编码元件，在低等真核生物和有些细菌中也发现有内含子(参见内含子、断裂基因)，**RNA 剪接**(RNA splicing)就是内含子精确去除。内含子根据剪接方式的不同，分为四种类型(表 27.2)。除了 tRNA 前体内含子，所有的剪接机制都涉及一对按顺序进行的转酯反应。在第一个反应中，一带有自由羟基的核苷酸进攻连接内含子和外显子的磷酸二酯键。这样释放了携带其本身羟基的 3′端外显子。在第二个反应中，自由羟基进攻连接内含子和下一个外显子的磷酸二酯键。第二个反应将外显子连接在一起，并释放内含子。三类转酯内含子有不同结构，起始羟基供体不同，形成中间体性质不同，在反应是否是自身催化或需要有反式剪接体提供等方面也存在差异。自身催化内含子称为**自我剪接内含子**，可以被认为是核酶(ribozymes，参阅)。

表 27.2 不同类型内含子和它们剪接方式的总结

内含子类型	内含子性质和剪接机制
转酯内含子——内含子中的剪接识别位点	
核 pre-mRNA 内含子	结构变化 自由羟基的起始供体由内部分支位点腺嘌呤提供 形成套索结构的中间体 需要装配大的反式作用的剪接体
Ⅰ型自我剪接内含子	保守的二级结构 鸟嘌呤核苷酸辅因子是自由羟基供体 没有套索结构形成 自我剪接
Ⅱ型自我剪接内含子	保守的二级结构 内部腺嘌呤是自由羟基供体 形成套索结构的中间体 自身催化剪接
Ⅲ型自我剪接内含子	与Ⅱ型内含子类似，但较小(100～200bp)包含数量有限的结构域
孪生内含子	多嵌合自我剪接内含子，常为Ⅱ型或Ⅱ型和Ⅲ型混合，经常以特定的顺序剪接。
非转酯内含子——剪接识别涉及外显子结构	
细胞核 tRNA 内含子	剪接机制类似于 tRNA 成熟——涉及剪切后的连接 没有中间体形成——内含子以线性片段切除 需要几种反式作用的加工酶

注：细胞核 mRNA 内含子和Ⅱ型自我剪接内含子使用非常类似的剪接机制，尽管后者是自身催化。

细胞核 pre-mRNA 内含子　　真核生物核基因组中的大多数内含子都不能自身催化剪接,称为**核 pre-mRNA 内含子**(nuclear pre-mRNA introns)。它们在大小和结构上有很多差异,但已鉴定了一些短的、保守的、控制剪接的顺式作用位点(参见图 27.1)。在这些剪接反应中,直接涉及**供体位点**(donor site,5′ site,left splicing site)、**受体位点**(acceptor site,3′ site, right splice site)和**分枝点**(branch site)。供体位点和受体位点具有位于保守程度稍差的序列中的高度保守的保守序列 GU 和 AG,称为 **GU-AG 原则**(GU-AG rule)。这些位点决定了内含子的边界,其改变或缺失的突变会破坏剪接(如不正常的剪接,见下文)。酵母中的分支位点具有 UACUAAC 的序列;但高等真核生物中除了提供起始羟基的腺嘌呤(在酵母序列中以斜体表示),此序列的保守性很差。

剪接反应以下面的方式进行(总结在图 27.1)。分支处活性的腺嘌呤的自由 2′羟基进攻连接上游外显子和内含子的磷酸二酯键。第一个转酯反应通过 5′→2′磷酸二酯键,连接内部的腺嘌呤与 5′剪接位点的鸟嘌呤,并形成套索结构。上游外显子 3′端羟基进攻连接内含子和下游外显子的磷酸二酯键。在 3′剪接位点进行的第二个转酯反应将两个外显子连接起来,并将内含子以**套索结构的中间体**(lariaf intermediate)形式释放出来,接着中间体被线性化(**去分支**)并降解。

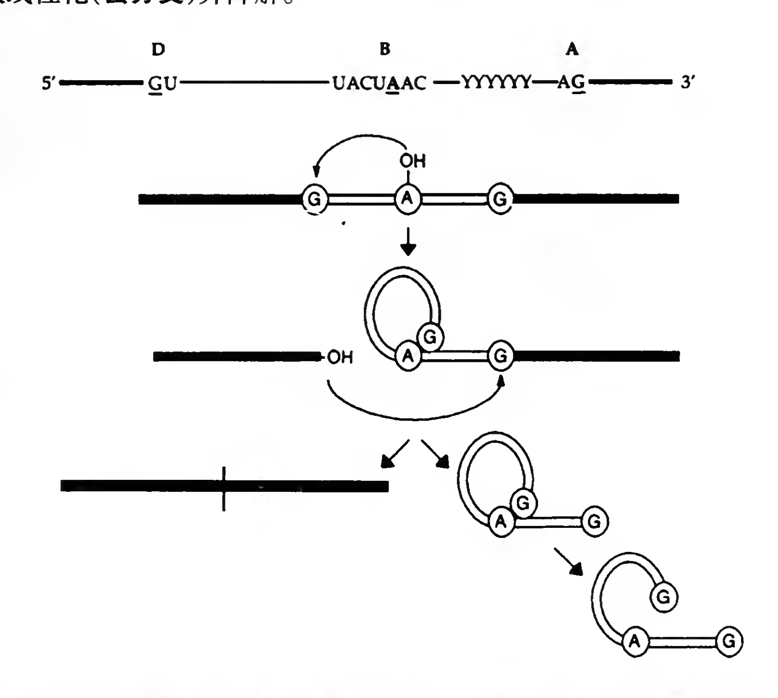


图 27.1　通过套索结构中间体的剪接。顶端的图显示了细胞核 pre-mRNA 内含子的典型结构,显示了供体(D),分支顺序(B)和受体(A)位点。在受体位点上游有多聚嘧啶区域,是剪接因子的识别位点。下面的图显示了剪接反应,内含子末端的鸟嘌呤和分支位点的活性腺嘌呤形成环状结构。Ⅱ型自我剪接内含子也遵循类似的途径。对细胞核内含子而言,反应由反式作用的剪接体催化,Ⅱ型内含子由 RNA 自身催化反应。

剪接体(spliceosome)是 40S～60S 的核糖核蛋白复合物,在 pre-mRNA 转录时形成,催化剪接反应。剪接体包含了**核内小分子核糖核酸蛋白**(small nuclear ribonucleoproteins)(snRNP,snurps),这种蛋白每个都由多个蛋白和一到两个**核内小分子 RNA(snRNA)**组

成。如上文所述，snRNA 常被称为 U-RNA，因为它含有修饰的尿嘧啶，分为 U1、U2 RNA 等。snRNP 是根据它所含有的 U-RNA 特殊种类而命名。那些在剪接中发挥作用的 snRNP 分别称为 U1、U2、U4/U6（含两个 snRNA）和 U5。每个 snRNP 都包含 10 到 20 个蛋白，其中有些是在所有的 snRNP 都存在。其他一些蛋白瞬时与剪接体结合，但其本身不是 snRNP 的成分，称为**剪接因子**（splicing factors）。剪接途径和剪接体的基本功能以及剪接因子在表 27.3 中总结。

核前体 mRNA 的剪接最显著的特征是在给定的种属信号下，大部分基因都能以精确的方式进行剪接。所有的剪接位点基本相同，这样理论上任何 5′位点可以与任何 3′位点连接。但在大多数有多个内含子的基因中通常不发生**外显子跳过**（exon skipping），只利用基本的剪接位点，而不是隐蔽的位点（参见下文的**选择性剪接**，alternative splicing）。对剪接中间体的研究揭示了很多基因的剪接途径。剪接位点的选择可能是转录本的内在性质，可能反映了剪接发生时的二级结构的变化（即剪接的顺序和专一性可能通过将非常规剪接位点放到二级结构中使其不能被利用来控制，一旦一个剪接成功地完成，构象的改变使进一步的剪接位点暴露出来）。精确识别的基础目前还不清楚，但有可能是通过称为 **SR 蛋白**（因为它们富含丝氨酸/精氨酸）的剪接因子调节，SR 蛋白与 U1 snRNP 和 U2AF 相互作用。这种相互作用可以通过两种方式调节：SR 蛋白铺展于内含子上，确定 RNA 的哪些部分需要被去除（**内含子桥连**，intron bridging）或者是铺展于外显子上，确定哪些部分的 RNA 需要保留（**外显子限定**，exon definition）。既然外显子比内含子相对要小，所以外显子限定的机制更为可能。U1 snRNP 刺激 U2AF 与上游 3′剪接位点的结合，支持了这种理论。

表 27.3 核 pre-mRNA 内含子的剪接——剪接途径和基本成分的功能

前剪接复合物	基本成分和事件
E 复合物（早期或定型复合物）	*U1* 和 *U2AF* U1 与供体位点结合 剪接因子 U2AF 与受体位点附近的嘧啶富集序列结合（这是下一步 U2 结合所需的）
A 复合物	*U1*、*U2* 和 *U2AF* U2 与分支点结合 需要 ATP 水解
B1 复合物	*U1*、*U2*，*U4/U6*、*U5* 和 *U2AF* U4/U6/U5 与前剪接复合物结合 U5 与 5′端的外显子作用，U4/U6 与 U2 结合
B2 复合物	*U2*，*U4/U6*、*U5* 和 *U2AF* U1 从前剪接复合物中释放出来 U5 再定位于内含子 U6 与供体位点结合 需要 ATP 水解
C1 复合物	*U2*、*U5*、*U6* 和 *U2AF* U4 释放 U5 再定位于受体位点 第一次转酯反应发生 需要 ATP 水解
C2 复合物	第二次转酯反应发生 剪接因子与套索结构中间体一起释放

剪接体在细胞中不是作为一个事先形成的复合物存在，而是在一定阶段被组装，称为**前剪接复合物前体**。

Ⅱ型自我剪接内含子　**Ⅱ型内含子**在植物和低等真核生物的细胞器基因组中发现。同

Ⅰ型内含子(见下文)一样,它们也能自身剪接,但与Ⅰ型内含子不同的是它们的剪接机制与核 mRNA 前体内含子使用的机制密切相关。剪接从内部腺嘌呤的 2′羟基进攻连接内含子与上游外显子的磷酸二酯键开始,产生套索结构和末端是 3′羟基的游离外显子。这个羟基在第二次转酯反应中进攻连接内含子和下游外显子的磷酸二酯键,使两个外显子连接并释放套索结构的中间体(参见图 27.1)。Ⅱ型自我剪接内含子的催化活性来自于其特征性的包含六个柄环结构域的二级结构,使外显子距离拉近。Ⅱ型自我剪接内含子的二级结构类似于核内含子的分支位点和 U2、U6 snRNAs,说明该内含子可能通过将催化功能由自身转移到反式作用的调节复合物,而由Ⅱ型自我剪接内含子进化而来。细胞核 mRNA 前体内含子在大小和结构上变化较大,而Ⅱ型自我剪接内含子保持均一性(内含子的进化的讨论参见第 22 章蛋白质:结构、功能和进化)。

Ⅰ型自我剪接内含子　Ⅰ型自我剪接内含子在线粒体基因组中发现,也存在于极少数单细胞真核生物(如嗜热四膜虫的 rRNA 基因中)的核基因组中。原核体系中少数内含子也是Ⅰ型内含子(如 T4 噬菌体胸苷酸合成酶基因)。Ⅰ型内含子是自身催化的,利用含鸟嘌呤的核苷酸辅助因子提供自由羟基。剪接反应按照如下步骤进行。辅助因子的 3′端自由羟基进攻连接内含子和上游外显子的磷酸二酯键,在内含子的 5′端加上鸟嘌呤使其延伸,并在上游外显子 3′端产生自由羟基。在第二个转酯反应中,上游外显子进攻连接内含子与下游外显子的磷酸二酯键,使外显子连接并将 5′有鸟嘌呤的内含子以线性形式释放(参见图 27.2)。这一片段可以通过第三个转酯反应,内含子的 3′端与鸟嘌呤连接形成环状结构。

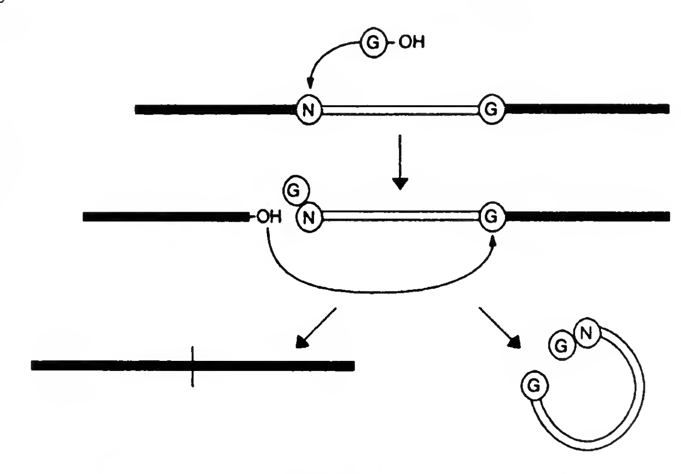


图 27.2　Ⅰ型内含子的自身剪接。详细的机制见正文。

反应的专一性由Ⅰ型内含子高度保守的二级结构所控制,Ⅰ型内含子有九个发夹结构,其中三个直接参与外显子的识别。特别是 P1 发夹结构包含了上游外显子的远端和内含子的起始碱基。这被称为**内部引导序列**(internal guide sequence, **IGS**),起先被认为是通过与两个外显子近侧区配对而使外显子并列,现在认为它决定了剪接的专一性。

内含子编码蛋白　一些Ⅰ型和Ⅱ型内含子含有编码促进剪接的蛋白的开放阅读框架。这些蛋白称为**成熟酶**(maturases),它们为剪接所需,但不具有催化活性(它们在结构上起

作用)。其他内含子编码蛋白的功能与内含子的迁移有关(参见第13章可移动的遗传因子)。Ⅰ型内含子经常编码内切酶,专一性剪切于DNA内含子插入位置,在修复介导的重组中转移内含子到没有内含子的等位基因上。这种过程称为**归巢**(homing),这种内含子称为**归巢内含子**(homing introns,参阅)。Ⅱ型内含子编码的蛋白有反转录酶和内切酶活性,使内含子象反转录子(retroposon)一样迁移。这样的过程称为**反转录归巢**(retro-homing)。也有些内含子编码的蛋白同时具有归巢和成熟酶活性。

在高等真核生物中,有些基因(含有其自身的内含子)以正向或反向完全嵌在一个大的基因内含子中(参见重叠基因、嵌套基因,T细胞受体基因)。内含子中也可能包含转录调节元件(参见增强子、沉默子)。

细胞核tRNA内含子　　细胞核tRNA基因中发现一类特殊的内含子。剪接不是通过顺序的转酯反应发生,而是剪切后进行连接。内含子通过特殊的内切酶反应从tRNA前体中被切除,产生5′羟基和3′环磷酸基团。两个末端分别产生适合连接反应的5′磷酸和3′羟基末端。修饰后的末端经过tRNA分子内碱基配对而并列,由连接酶作用而连接在一起。tRNA内含子的其他性质也不同寻常,它在结构或序列上没有保守性,也就是对内含子的识别主要是通过外显子结构。在酵母中,尽管不同tRNA基因的内含子没有关系,它们经常在相同的位置被发现——在反密码子的下一个碱基。内含子有与反密码子互补的序列,使它们能够与反密码子环配对。这样的二级结构可能是正确剪接所需的:破坏二级结构的突变使剪接效率下降。

通过剪接产生的变化　　对很多基因而言,RNA剪接是一个不变的加工过程。基因转录时所有细胞中产生同样的成熟转录本及单种产物。这称为**组成型剪接**(constructive splicing)。

其他一些基因,RNA剪接可以作为基因调节的机制,也就是说在两种组织中同一基因表达不同,尽管它以相同的方式和速度转录。这样的调节以两种方式发生。

在第一种情况,**加工**或**去除调节**(processing or discard regulation),初始转录本在一些组织中加工,在其他组织中不剪接。在有些低等真核生物中(如海胆),这是主要调节机制。大部分基因在大部分组织中转录,由RNA加工调节来决定那些基因被翻译。在果蝇中,一个类似的机制用来控制P因子转座酶(参阅)的合成,在这种情况下虽然只有一个内含子未被剪接,但却限制了P因子转座只发生在生殖细胞中。转座酶基因包含三个内含子,在生殖细胞中产生成熟的编码转座酶的转录本,然而在其他组织中,第三个含终止密码子的内含子被保留,使转座酶蛋白截短,从而阻止体细胞中P因子的转座。内源性基因也用同样的策略,如哺乳动物$GABA_A$受体ε亚基的mRNA前体除了脑以外其他组织都不剪接。

在第二种情况,**差别**或**选择性**剪接(differential or alternative splicing),基本转录体通过选择使用外显子以不同方式加工。相关成熟转录体家族产生不同基因产物,称为**剪接变异体**(splice variants)或**剪接异构体**(splice isoforms)。选择性剪接以多种方式进行(图27.3)。

(1)通过选择启动子删去5′外显子;

(2)通过多腺苷酸位点选择删去3′外显子(上两种情况都是转录水平的调节:不同的

RNA 加工是选择使用转录控制元件的二级结果)；

(3)在外显子确定中对外显子 5′或 3′剪接位点的选择；

(4)对外显子 5′或 3′剪接位点的选择，跳过一个外显子；

(5)外显子的互斥效应(mutual exclusion)。

这些过程都是 RNA 加工水平的调节，涉及的因子在下文中讨论。

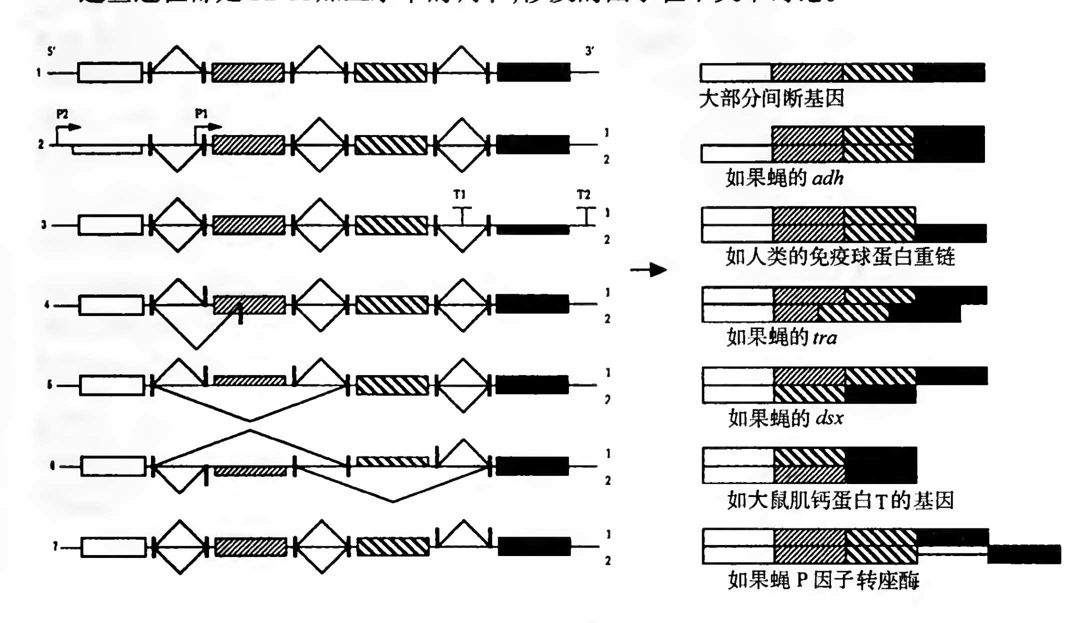


图 27.3 选择性剪接的形式。初始转录本和剪接途径在图左表示，成熟的转录本在图右显示。外显子以方框表示，内含子以水平线表示。折线表示剪接反应，不同的剪接显示在初始转录本的上下，不同的产物在相应的位置。(1)组成型剪接；(2)果蝇乙醇脱氢酶基因选择不同的启动子产生不同的基因产物；(3)免疫球蛋白使用不同的多腺苷酸化位点产生分泌和跨膜形式的产物；(4)果蝇性别转换基因(transformer)mRNA 选择不同的 3′剪接位点产生有功能和没有功能的产物；(5)果蝇两性基因(doublesex)mRNA 外显子的跳过产生不同的产物；(6)大鼠肌钙蛋白 T(troponin T)基因外显子 16 和 17 的互斥效应产生了不同类型的产物；(7)果蝇 P 因子转座酶因一个内含子能否剪接，产生了有功能和无功能两类 mRNA，此可被看作一种单个内含子加工的调节或丢失调节。

选择性剪接的控制　如上文中所讨论的，定向剪接是由 U1 snRNP 和剪接因子 U2AF 以及与其相互作用的 SR 蛋白起始的。选择性剪接机制将这些剪接因子导入前体 mRNA 的剪接位点上，利用特殊的剪接位点或通过刺激，或通过抑制来起作用。有两种不同剪接控制方式：一种持续作用，取决于特殊剪接因子的量；第二种以组织特异的方式作用，涉及特殊的剪接控制因子。

数量控制机制常发生在利用 5′和 3′位点作不同的剪接。例如对 SV40 T/t 抗原基因的差异剪接分析发现了一种选择剪接因子(ASF)，是与组成型剪接因子 SF2 类似的 SR 蛋白。结果显示，当存在选择性的 5′剪接位点时，高浓度的 ASF/SF2 倾向于使用近侧位点，低浓度则倾向于使用远侧位点。第二种因子，hnRNPA1 拮抗 SF2 的作用。不同细胞中这两个因子相对量的不同可以解释细胞类型专一性的剪接异构体。

当剪切异构体以严格的细胞类型特异的方式产生时，数量模型不能完全解释。在真核生物中有很多这类的例子，在果蝇中，控制性别决定的剪接分阶段调节可能是最明显的

选择性调节的例子(图 27.4)。组织专一性的剪接因子通过类似转录调控的顺式作用元件[**剪接增强子**(splice enhancers)和**剪接阻遏物**(splice repressors)]促进或抑制特殊的剪接信号。在果蝇中用遗传分析发现了组织专一性的剪接调节因子,更多的这类机制在哺乳动物中被发现,包括限于心脏和脑的蛋白 SmN。

特殊的剪接途径　　异常剪接(aberrant splicing)可以是由于野生型基因剪接位点突变或新的剪接位点的产生而引起的。**隐蔽的剪接位点**是弱的共有位点,通过正常位点的缺失

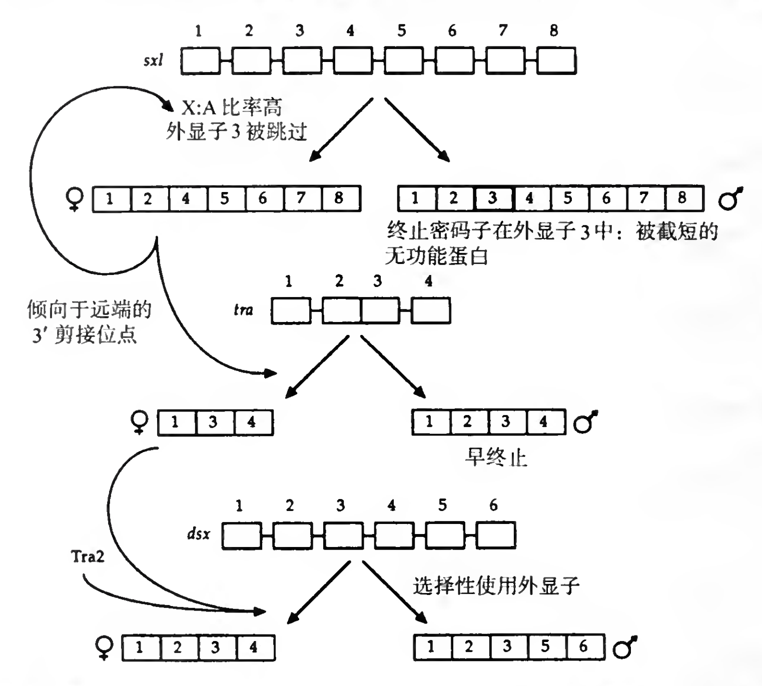


图 27.4　果蝇的性别决定。性别致死基因(*sxl*)外显子 3 的不同剪接在雌性中产生活性蛋白而在雄性中是截短的没有功能的蛋白。Sxl 是促进本身基因和下游性别转换(*tra*)基因剪接的剪接因子。在 Sxl 的影响下,*tra* 的 mRNA 前体剪接倾向于远端的 3′剪接位点,去除了会引起功能失活的 2 号外显子,Tra 也是剪接因子,与另一剪接因子 Tra2 一起改变了两性基因(*dsx*)雌性专一性的剪接途径,启动雌性专一性的分化。Sxl 产生依赖于果蝇中性别决定的平衡机制(参见第 6 章发育,分子机制)。

显露出来。几个隐蔽位点经常同时出现,产生多种异常的外显子序列缺失或内含子保留的成熟转录本。很多形式的地中海贫血可追踪为剪接位点突变而产生异常球蛋白的转录本(参见文框 15.1)。

另一种形式的特殊剪接是**反式剪接**(trans-splicing),使不同 RNA 的外显子被连接在一起,即在两个或更多不同的基因的初始转录本(通常剪接应该是**顺式剪接**)之间发生。反式剪接发生在一些叶绿体基因组中,分隔的基因(参阅)以不同方向,分散排列的外显子的形式存在(参见细胞器基因组、基因)。反式剪接也发生在低等真核生物的细胞核基因组中:锥虫中提供了一个极端的例子,同样 35 个碱基的**拼接前导 RNA**(spliced leader RNA, **SL RNA**)通过加工加在所有的 mRNA 上。线虫中有 22 碱基的 SL RNA 加到一些 mRNA 上,包括三个肌动蛋白。这种情况下,可选择的反式剪接使多顺反子的 mRNA 前体可以产生多种形式的可翻译成熟转录本(参见操纵子)。通常,反式剪接的引导部分有 5′剪接位点而没有 3′受体位点,3′端的基因有上游不配对的受体位点。SL RNA 拥有三

个茎环和单链区的二级结构，被认为与 U1 snRNP 以同样的方式起作用。感冒病毒的抢帽(参阅)也是反式剪接的一种形式。

27.4 RNA 编辑

RNA 编辑　在大多数情况下，编码蛋白基因的信息直到翻译并不改变。基因可以被那些将被剪接去除的内含子所打断，基因的信息可以被选择性的使用，信息的确切意义没有改变。**RNA 编辑**(RNA editing)是改变 mRNA 外显子信息的协同转录或转录后的机制。这过程主要限于细胞器基因组，虽然也有几个哺乳动物细胞核 mRNA 的局部编辑的例子。RNA 编辑产生的基因称为**隐蔽基因**(cryptogenes)：基因产物的结构不能从基因组 DNA 序列中推导得到。RNA 编辑在进化中的作用还不清楚。嘧啶插入是主要的编辑过程，因而编辑可能是将嘧啶导入嘌呤丰富序列的一种机制。编辑能够产生重要的功能性结果，如在哺乳动物中它决定了一些离子通道和 G 蛋白偶联受体的性质。主要有四种类型的 RNA 编辑(表 27.4)。

表 27.4　四种 RNA 编辑机制

编辑过程	性质	举例	机制
简单编辑	单碱基转变的转录后调节	C→U 转换，哺乳动物载酯蛋白 A→G 转换，哺乳动物谷氨酸受体亚基和 5-羟色氨受体。编辑也发生在内含子 C→U 转换，在植物细胞器中	特殊的胞嘧啶脱氨酶修饰 由双链 RNA 脱氨酶将腺嘌呤转变成次黄嘌呤。发现三个基因 不清楚
插入编辑	插入单个核苷酸或少量核苷酸的丢失，协同转录	副粘病毒 P 基因转录时 G 的插入	转录链的跳格
泛编辑	插入/缺失多个尿嘧啶核苷酸，转录后	U 插入/缺失，椎虫动基体 mRNA(参见动机体 DNA)	编辑序列由外源反义引导 **RNA (gRNA)** 提供，gRNA 在**编辑体**(editosome)核蛋白颗粒中与前编辑 mRNA 配对，鉴别作为错配的位点进行编辑。gRNA 有多尿嘧啶尾巴，提供以 3′→5′极性插入的尿嘧啶
	插入多个胞嘧啶	C 插入至少在四种 *Phisarum polycephalum* 的线粒体 mRNA 中	不清楚
多聚腺嘌呤编辑	在转录本末端加腺嘌呤完善终止密码子	脊椎动物线粒体 mRNA 的多聚腺苷酸化	缺少终止密码子的前体 mRNA 被多聚腺苷酸化，由第一或第二个腺嘌呤提供缺少的信息

泛编辑详细情况可参见动肌体 DNA。

27.5 后加工调节

RNA 转运和亚细胞定位　真核生物中 mRNA 在细胞核合成，必须转运到胞浆中进行翻译。人们认为 RNA 加工复合物被限制在细胞核内对于下一步将 RNA 运输出核孔起到重要作用。运输的具体机制还未完全研究清楚。目前知道是有选择地转运 mRNP，并

需要 ATP 水解。5′端帽子对转运很重要(不加帽的转录本,如 rRNA 和 U6 snRNA 不被转运),剪接体成分的存在阻断转运(这样可以避免对部分被剪接的转录本进行翻译)。在转运前,一些 hnRNP 蛋白从转录本上被切除,另一些在转运过程中解离下来,返回细胞核。加工后的 mRNA 离开细胞核后立即与核糖体结合。

在真核生物中,RNA 从细胞核转运过程是潜在的调节对象,但到目前为止只有病毒 RNA 是以这种方式调控。研究的最清楚的是 HIV 基因组,它的 RNA 剪接和转运由 **Rev 蛋白**和内含子中顺式作用元件(Rev 相应元件)调控。Rev 似乎帮助结合有剪接体的 RNA 转运,并允许部分剪接的 HIV RNA 转运。目前还不清楚 Rev 蛋白是如何避免正常细胞对这种转运的抑制作用(参见文框 30.3)。

一旦进入细胞质,mRNA 与核糖体结合就可以自由扩散或被定位于细胞内某一特殊区域。部分被翻译的编码分泌蛋白的转录本经常运输到粗面内质网的膜附近,这样多肽可以被易位入这种细胞器的腔内(参见信号肽)。这依赖于多肽的定位信号或 RNA 本身所携带的信号(**RNA 导向**,RNA targeting)。某些转录本与细胞骨架相缔合,定位于细胞的特殊区域。后者在动物发育过程是一个普遍的现象,它提供了在卵细胞中作为细胞质决定(参阅)的定位信号的机制(如果蝇发育中的 *bicoid* 和 *nanos* mRNA;非洲爪蟾发育中的 *vg-1* 的 mRNA)。在其他情况下,定位直接与功能有关(如成肌细胞分化中 α-肌动蛋白和 β-肌动蛋白的 mRNA 定位)或对细胞的生存至关重要(如髓磷脂基本蛋白翻译的错误定位是致死的)。在所有情况下,定位信号都被发现在转录本的 3′非翻译区,并与一个或多个定位所需蛋白结合。

mRNA 更新　特定转录本的丰度决定于它的合成速度和稳定性,稳定性反映了降解速度。mRNA 的稳定性决定了当转录的速度改变时,稳态的 mRNA 的变化情况以及提供用于蛋白合成的量。如同其他基因调节一样,RNA 的稳定性可以是不变的也可以是被调节的。**RNA 的半衰期**(mRNA half-life)也就是 mRNA 降解速度的变化,可以快速和瞬时改变特定 mRNA 的量而不改变其转录活性。

细菌中的 mRNA 更新和反向调节　细菌 mRNA 的半衰期较短,大多数稳定的转录本是几分钟,这样可以通过调节转录速度使蛋白合成速度对环境作出快速反应。原核生物 mRNA 的降解由 RNA 内切酶(也称为 **RNA 限制性内切酶**)和 3′→5′外切酶所调节。mRNA的二级结构是决定稳定性的主要因素,大多数稳定的转录本在 3′端非翻译区拥有多个发夹和茎环结构,保护其免受外切酶作用。含有内切酶作用位点的转录本特别不稳定。尽管破坏大肠杆菌的核糖核酸酶 E 使 RNA 稳定性增加 2～3 倍,但细菌中降解 RNA 的酶目前还不得而知。

细菌中转录、蛋白合成和降解的紧密关联使其能够通过称为**反向调节**(retroregulation)的特殊方式进行基因表达调节,RNA 降解在转录终止水平上进行调节。基因表达取决于转录的基因下游是否存在顺式作用元件。如果存在,新生的 RNA 的结构有利于快速降解,阻断翻译的进行。如果转录在这一位点前终止,RNA 就相对稳定,蛋白合成得以进行。例如 λ 噬菌体就用反向调节控制整合酶基因的表达(参见文框 30.1)。

真核生物 mRNA 的更新　真核生物 mRNA 比细菌转录本稳定得多。酵母 mRNA 半衰期约在 5 到 45 分钟,后生动物则更稳定,平均在 10 小时左右,反映了多细胞生物细胞的

环境是相对稳定的。哺乳动物 mRNA 多聚腺苷酸尾巴和维持尾巴长度的 PABP 结合与稳定性有关。去腺苷酸化或 PABP 减少时哺乳动物细胞 mRNA 快速降解,但在酵母中,PABP 的出现是降解的信号,所以它精确的作用还不清楚。真核生物中外切酶对多聚腺苷酸尾巴的降解是 mRNA 降解的第一步。没有多聚腺苷酸尾巴的组蛋白 mRNA 为特殊的内切酶降解。有证据表明真核细胞中存在内切酶活性。

一些真核生物的调节蛋白需要瞬时爆发活性,所以它们的 mRNA 与细菌转录本一样不稳定。这包括很多立即早期基因的转录本,这些基因被信号传导通路诱导,需要产生短期的调节反应(参见 c-*fos*,c-*jun*;见细胞周期,癌基因和癌症,信号传导各章)。很多真核生物不稳定的 mRNA 都有特殊的不稳定元件,经常是 **AU 丰富元件(AUREs,AREs)**,如多拷贝的 AUUUA 序列,一般在 3′非翻译区(如白介素 1,干扰素 β 和 c-*fos* 的 mRNA),也有的位于编码区(如 β 微管蛋白和 c-*myc* 的 mRNA)。尽管这些不稳定性所依赖的机制还不清楚,AUREs 有助于形成茎环结构的能力说明影响 RNA 折叠的因子可能调节 RNA 的稳定性。不稳定元件看来像是相对独立的,因为当它们插入异源 mRNA 的 3′非翻译区时会引起不稳定。稳定性可能取决于二级结构或上下游序列,很多稳定的转录本中发现了相似的元件,如神经元专一性的烯醇化酶 mRNA。

少数真核生物 mRNA 的稳定性可能受到与不稳定元件结合的反式作用因子的调节。在细胞内离子水平过量的情况下,运铁蛋白的受体 mRNA 发生降解。据推测可能是按照上述组成型不稳定转录本降解调节相同的过程进行。当离子水平下降,一个因子结合在转录本**铁效应元件**(iron response element, IRE)上,并阻止其降解。IREs 是 30 碱基的基序,形成含有不稳定元件的茎环结构。IRE 结合蛋白的存在被认为阻断了不稳定元件。在其他转录本中,同样的 IRE 用来调节翻译(见第 23 章蛋白质合成)。

(陈 淳 译)

第28章
信 号 传 导

基本概念和定义

- 细胞通过重新组织它们的结构、调节蛋白活性和改变基因表达模式来对环境作出反应。这种对反应的刺激称为**信号**(signal),信号可以是小分子、大分子或物理试剂,如光。信号通过称为**受体**(receptors)的分子与反应的细胞作用。
- 小分子经常以可扩散的信号起作用。在单细胞生物体中,可扩散的信号可以是起源于环境或由其他细胞释放的(如酵母的交配型信息素、网柄菌属(*Dictyostelium*)的cAMP)。在后生动物中,信号可以从邻近的细胞中释放,在短距离上传播(**旁分泌信号**,paracrine signaling),或者由远处的细胞分泌,通过血管系统达到它们的目标(**内分泌信号**,endocrine signaling)。大分子信号经常与细胞外基质相联系,或出现在邻近细胞的表面(**近分泌信号**,juxtacrine signaling)。一个与受体结合的信号分子称为**配体**(ligand)。
- 信号可以通过三种途径起作用。某些化学信号可以穿过细胞质膜与内部受体相互作用(如甾类、氧化氮)。大多数信号是保留在细胞外的亲水分子。它们与跨膜(穿膜的)或膜相关受体作用,引起受体结构的改变。这样的相互作用产生**信号传输**,也就是说信号分子内化(受体构象改变或产生一个小孔引起的运输,如在离子通道的情况中或通过受体介导的内吞)。此外,受体构象的改变可以诱导细胞内酶的活性,介导下游效应,而配体留在细胞的外部(**信号传导**,signal transduction)。物理刺激也可以与受体作用或直接调节它们的效应。当光子引起光敏感分子11-顺式视黄醛改变为全反式构象,光刺激G蛋白与视紫红质和视锥蛋白受体连接。相反,对热休克和类似刺激的反应是通过细胞内变性蛋白的增加直接介导的。
- 信号传导涉及顺序酶激活和称为第二信使的小分子水平的调节。这样使原始信号得以放大(直接扩散和信号传输只是线性反应)。信号传导途径可以是**会聚的**或**分散的**,使多种刺激产生类似的反应或单个信号产生不同的反应。进一步的分散由对刺激长度和强度的不同反应,细胞类型特异合成的不同受体和信号分子而产生。信号传导途径由复杂的调节网络控制,对刺激的反应决定于细胞相反力量之间的平衡。
- 信号的交付涉及对转录因子、酶和细胞结构成分的激活或抑制,这经常是通过改变它们的磷酸化状态来达到。

28.1 受体和信号途径

细胞表面的受体和它们的配体　细胞对广泛的信号反应是通过同样广泛的受体进行。在后生动物中，很多信号是小的多肽。短距离（旁分泌）作用的多肽信号称为**生长因子**(growth factor)（或在造血系统中为**细胞因子**），它们可根据结构或功能的类似性分为不同的家族。这些分子不仅与细胞生长有关，而且与分化、迁移和其他细胞功能有关。**激素**是分泌信号分子，其中很多是类似生长因子的多肽，但也有大的糖蛋白激素和类固醇激素，后者能够直接通过细胞质膜扩散，与细胞质或细胞核内的受体作用。另一类重要的多肽信号分子是**神经肽**(neuropeptides)，它们在中枢和外周神经系统介导神经传导和神经调制。细胞可以对很多小分子，如氨基酸、核苷酸和生物活性胺，和细胞外基质或邻近细胞表面控制细胞运动以及黏附的大分子发生反应。一些受体与邻近细胞的**反受体**(contra-receptors)结合，这样细胞间的相互作用诱导了交互信号。

受体多样性比它们的配体要小一些，因为细胞反应的种类比刺激少一些。这样很多刺激可产生同样类型的反应（生长抑制、生长、特定基因的转录），这是通过将很多信号归入共同的途径而产生的。这样，任何类型受体都有配体结合多样性结构域，但是受体作为一个整体可以归入少数有共同效应结构和信号机制的家族。目前对研究人员来说一个问题是为何有如此之多的看来是起同样作用的受体亚型。几种常见受体家族的结构和活性在下面将作讨论。

离子通道　离子通道是细胞膜上水溶性的小孔，它们的活性通过开放和关闭控制，从而允许或阻止离子或其他小分子进出细胞。离子通道在缺乏状态时是关闭的，对特殊信号反应后开放。有些是**配体门控**(ligand-gated)，也就是说它们对特定配体结合反应而开放〔如谷氨酰胺和 γ-氨基丁酸(GABA)受体〕。其他受体特别是在神经元中发现的是**电压门控**(voltage-gated)，也就是说对与电位作用相关的电流改变作出反应而开放。还有一些是**第二信使门控**(second-messenger-gated)的离子通道，对细胞内的第二信使，如钙离子、环核苷酸和脂（见下面的*第二信使*）。离子通道允许离子沿着电化学梯度正向移动或通过活性离子泵逆梯度运动，后者需要 ATP 水解。

离子通道的跨膜区域包含几个*疏水螺旋*（参阅），堆积在一起使疏水残基包围了亲水残基。有些离子通道是多个类似结构域组成的单个大分子。另一些是多亚基蛋白。神经传递-门控离子通道已被细致地分析，包含了五个类似的亚基，其中每个都有四个螺旋的跨膜结构域。四个螺旋中的一个是双亲性的，通道的小孔是由每一亚基的双亲性的螺旋排列而成。离子通道的开放和关闭由门控机制（如配体结合或电压改变）刺激产生构象变化来调节。小孔进口处带电残基的特定排列控制了通道的离子选择性。

G-蛋白-偶联受体　**G 蛋白偶联受体(GPCRs)**是中心疏水区域形成七次跨膜结构域的单个多肽。在外部，N-端结构域经常是配体结合结构域，尽管有些配体与膜结构域结合。在内部，C 端区域与三聚体鸟嘌呤核苷酸结合蛋白(**G 蛋白**)相结合。数百个 GPCRs 的基因被克隆和分析。证明这些受体对应它们配体的多样性的显著保守性，它们的配体包括小分子物质，如肾上腺素和 5-羟色胺，如胰高血糖素、神经激肽、阿片样物质的多肽和像

促卵泡激素的大分子糖蛋白。进而有些 GPCPs 可以被增味剂和光线刺激。GPCPs 也是广泛分布的——有上百种脊椎动物受体被发现,GPCPs 与酵母交配型信号传导有关,对网柄菌属的 cAMP 作出反应。在病毒基因组中发现了两个公认的 GPCPs,尽管它们的功能还不知道。在盐细菌(*Halobacterium balobium*)中也发现了结构类似但顺序不同源的膜蛋白。

G 蛋白是异源三聚体复合物(包含 α、β 和 γ 亚基),在非活性形式时与 GDP 结合。被配体结合激活的受体加快了 GDP 从 G 蛋白中解离的速度。解离使 GTP 替代了 GDP,因为它在细胞中含量比 GDP 更多,这引起 α 亚基从 βγ 二聚体中解离出来。G 蛋白有内在的 GTPase 酶活性,缓慢水解 GTP,将本身灭活,重新组建成三聚体。

在哺乳动物中已经发现了 20 多种类型的 α 亚基和多种类型的 β 和 γ 亚基:可以形成上百种组合的三聚体,有些可以在特定细胞类型中共表达。α 亚基经常是下游效应分子的基本激活因子,发现的四类主要的 G 蛋白家族的 α 亚基与不同类型的效应分子作用,对细胞内第二信使产生不同的效应(表 28.1)。βγ 二聚体也可以介导下游反应,如酿酒酵母的 STE1 和 STE2 蛋白激活联系 G 蛋白信号到 MAP 激酶途径的激酶(见下)。对 G 蛋白激活的精确反应由特定的 αβ 和 γ 亚型和下游靶分子的特殊亚型同时控制。如不同类型的腺苷环化酶显示了对 G_s 和 G_i 调节的不同反应。

表 28.1 G-蛋白 α-亚基家族,主要亚型,它们的效应分子和被激活的第二信使

家族	主要亚型	效应分子功能	第二信使的反应	举例
G_s	G_s	刺激腺苷环化酶和 Ca^{2+} 通道	cAMP 增加,Ca^{2+} 通道开放	组胺(H2 受体)
	G_{olf}	刺激腺苷环化酶	cAMP 增加,	嗅觉受体
G_i	G_{i1-3},G_0	抑制腺苷环化酶	cAMP 减少	δ 阿片样受体
		刺激 K^+ 通道	膜电位改变	α2 肾上腺素能受体
		抑制 Ca^{2+} 通道	Ca^{2+} 通道关闭	
	G_t	刺激 cGMP-PDE	cGMP 减少	视紫红质
	G_{gust},G_z	不清楚	不清楚	味觉受体
G_q	G_q,G_{11},$G_{14\sim16}$	激活磷脂酶 C-β	Ins(1,4,5)P_3 增加,DAG 增加	速激肽受体 有些谷氨酸受体
G_{12}	G_{12} G_{13}	刺激/抑制 Na^+/H^+ 交换	pH 改变	

缩写:(ACG)MP,环腺(鸟)苷酸;PDE,磷酸二酯酶;Ins(1,4,5)肌醇三磷酸;DAG,二酰甘油。

酪氨酸受体激酶　　酪氨酸受体激酶(**RTKs**)是一次跨膜受体,在细胞质结构域中有内在的配体激活蛋白酪氨酸激酶活性。已发现多于 50 种 RTKs。根据它们细胞外结构域的结构基序分为 14 个主要家族,尽管在 PDGFR,FGFR 和 VEGFR 亚家族中结构域是根据结合其他信号分子的内部顺序分类的。

在没有配体的情况下,RTKs 是单体且没有激酶活性。尽管不是所有受体激活机制都被阐明,激活可能是由于多聚化引起的,多聚化刺激了激酶活性和**自身转磷酸化作用**(autotransphosphorylation)(一个受体单体被另一个磷酸化)。很多 RTKs 在二聚体是有活性的,二聚体化是由配体结合引起的,经常是因为配体本身是二聚体,可以同时结合两个受体(如 PDGF)。胰岛素受体家族在这方面是相反的,它们的受体始终是二聚体,两个亚基通过二硫键结合在一起。在这种情况下,配体结合可能是诱导了有利于激酶结构域激活的构象改变。

一旦受体被激活，RTK 信号可以由两种途径起始：①磷酸化下游靶分子和②重建包含特异性识别磷酸化残基的蛋白质的信号复合物。其他蛋白的几个结构域与磷酸化酪氨酸特异性作用，包括 SH2 结构域和 PTB 结构域（见下文）。胰岛素受体家族又与其他 RTKs 不同，它的下游反应是通过小蛋白，胰岛素受体底物－1(IRS-1)被胰岛素受体激酶磷酸化而介导。通过 RTK 激活起始的主要信号途径包括 Ras-Raf-MAP 激酶途径和磷脂酶 C-γ-激活的第二信使体系，这些将在下面讨论。

与酪氨酸激酶活性相关的受体　　很多受体磷酸化激活靶蛋白，但没有内在的酶活性。这样的受体，包括很多细胞因子受体和免疫系统中加工抗原的受体，在配体结合后征集和激活细胞质酪氨酸蛋白激酶。

有两类主要的细胞因子受体家族。Ⅰ型受体包括大多数造血和免疫系统的细胞因子和三个普遍存在的受体：gp130，βc 和 γc。关系较远的Ⅱ型受体包括干扰素受体和白介素 10 受体。受体的不同类型是根据分子结构域在细胞内外的保守基序决定的(图 28.1)。

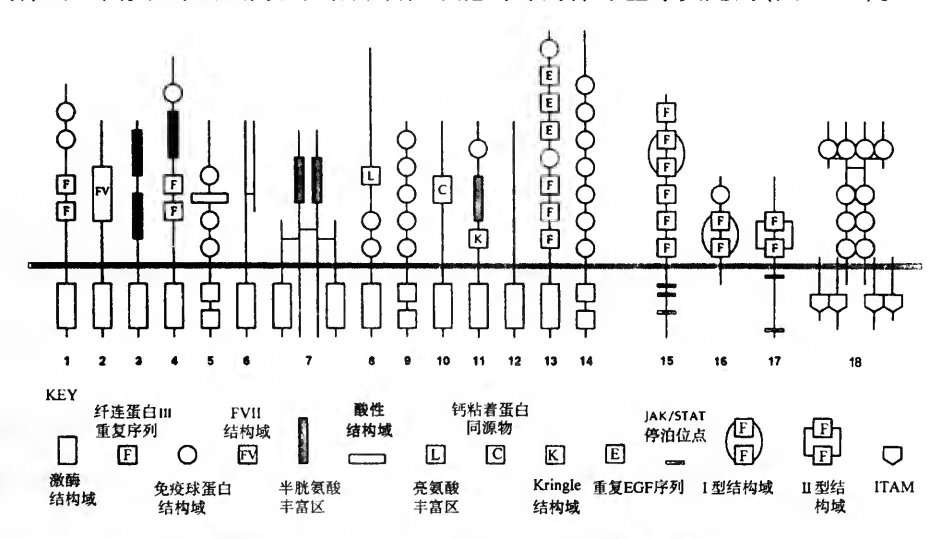


图 28.1　有内在的或与酪氨酸激酶活性相关的受体结构。1～14 是受体酪氨酸激酶的不同亚家族的例子。1，ARK；2，DDR；3，EGFR(表皮生长因子受体)；4，Eph；5，FGFR(成纤维生长因子受体)；6，HGFR；7，胰岛素受体；8，NGFR(神经生长因子受体 TrkA)；9，PDGFR(血小板衍生生长因子受体)；10，c-RET；11，ROR1；12，c-Ryk；13，TIE；14，VEGFR。15～17 是不同细胞因子受体亚家族的例子。15，gp130；16，IL-6R(白介素－6 受体)。这些是Ⅰ型受体。17，IFNAR1(干扰素－α 受体)，一个Ⅱ型受体；18，免疫球蛋白 B 细胞受体。

配体结合诱导受体二聚化，将两个单体的细胞内结构域并列。很多Ⅰ型受体以同源二聚体发挥功能，其他与普遍存在的 p130，βc 和 γc 受体形成异源二聚体。异源二聚化是非混杂的，所以造血细胞因子受体可根据所作用的普遍存在的分子分成三个亚家族。在每个亚家族中，有一定程度的功能冗余，反映了共同的受体单体。Ⅱ型受体是多亚基组成的，其中很多未被鉴定。Ⅱ型受体的激活涉及多聚化。

每个受体的细胞内结构域始终与 Janus 家族的蛋白酪氨酸激酶相联系(JAKs；表 28.2)。二聚化刺激 JAKs 交互的自身转磷酸化作用，JAKs 被激活并磷酸化其受体本身。受体的磷酸化酪氨酸残基通过 SH2 结构域征集 STATs(信号传导和转录活化蛋白)(表 28.2)。STATs 的磷酸化诱导二聚化，并转入细胞核，作为转录调节因子起作用。

表 28.2 Ⅰ型和Ⅱ型细胞因子受体的信号传导成分

受体复合物	募集基序	JAKs	STATs
Ⅰ型受体			
EPOR	XXYLVL/V	Jak2	STAT5
G-CSFR		Jak1, Jak2	STAT3
GHR		Jak2	STAT1
GM-CSFR(βc)		Jak1, Jak2	STAT5
IL-2Rb(γc)	DAYLSL, DAYCTF	Jak1, Jak3(也是Lck, Syk)	STAT5
IL-3R(βc)		Jak1, Jak2	STAT5
IL-4R(γc)	XGYKPFG, GYKAFS	Jak1, Jak3	STAT6
IL-5R(βc)		Jak1, Jak2	STAT5
IL-6R(gp130)	XXYXPQX XXYXXQ	Jak1, Jak2, Tyk2	STAT1 STAT3
IL-7R(γc)		Jak1, Jak3	STAT5
LIFR(gp130)		Jak1, Jak2, Tyk1(也是Yes, Hck)	STAT3≫STAT1
Ⅱ型受体			
IFANR1	XXYXXQ	Tyk2	STAT1, STAT2
IL-10R		Jak1, Tyk2	STAT1
INFAR2		Jak1	STAT1, STAT3
INFGR		Jak1, Jak2	STAT1, STAT3

Ⅰ型受体与普遍存在的受体 gp130、βc 和 γc 之一二聚化，特定的共同成分在括弧中表示；其他则以同源二聚体起作用。有些细胞因子受体与不属于 Janus 家族的酪氨酸激酶作用——其他激酶也在括弧中表示出来。对有些受体，征集 STATs 的磷酸化酪氨酸基序通过突变被发现。

缩写：EPOR，促红细胞受体；G-CSFR，粒细胞集落刺激因子受体；GHR，生长激素受体；GM-CSFR，粒巨噬细胞集落刺激因子受体；IL，白介素；LIFR，白血病抑制因子受体；IFNA(G)R，干扰素 α(γ)受体；JAK，Janus 激酶；STAT，转录信号传导和激活因子。

抗原受体是独特的，因为通过体细胞重组[参见 V(D)J 重组]和在免疫球蛋白中的体细胞高度突变(参阅)产生独特型的分布。它们是多成分受体，有不同结构的配体结合亚基和不同的征集激活信号复合物的亚基。所有抗原受体的不同亚基都有称为 **ITAM(免疫酪氨酸激活基序**，见图 28.1)，它作用并激活细胞质蛋白酪氨酸激酶。已经有很多非受体的酪氨酸激酶被发现(图 28.3)，Lck、Fyn 和 Zap 对 T 细胞受体信号很重要，Blk、Fyn、Lyn、Syk 和 Tec 家族的成员对 B 细胞受体信号重要，Csk 激酶起明显的调节作用。

受体丝氨酸/苏氨酸激酶　　信号分子的 TGF-β 超家族成员受体(表 28.3)根据大小和配体结合特异性分类。很多Ⅰ型和Ⅱ型受体被发现有内在的丝氨酸/苏氨酸激酶活性。Ⅰ型受体特征性拥有比Ⅱ型受体小的半胱氨酸富集的细胞外结构域和 C 端尾巴，并在激酶结构域附近有一独特的 GS 结构域(甘氨酸和丝氨酸富集)。

涉及受体丝氨酸/苏氨酸激酶的信号传导需要受体的多聚化。TGF-β 先开始与持续激活的Ⅱ型 TGF-β 受体结合，结合后征集了Ⅰ型 TGF-β 受体，它可在丝氨酸残基上通过磷酸化而被激活，它本身再磷酸化下游的靶分子。超家族中的其他信号分子可直接与Ⅰ型受体结合，特别是骨形态发生蛋白。

几个 TGF-β 信号的下游靶分子是细胞周期依赖蛋白激酶的抑制蛋白(见第 2 章细胞

表 28.3　生长因子 TGF-β 超家族,它们的Ⅰ型和Ⅱ型受体和已知的下游靶分子

家族	配体	Ⅰ型受体	Ⅱ型受体	已知的目标蛋白
TGF-β	TGF-β1/5 TGF-β-2 TGF-β-3	TβR-Ⅰ(Alk-5)	TβR-Ⅱ	Smad2(Smad3)
活化素/抑制素	活化素 几种抑制素	ActR-Ⅰ, Act-IB, Atr1	Act-RⅡ, ActRIIB Punt(Atr-Ⅱ)	
BMP(DPP)	Dpp/ BMP-2 BMP-4	Punt(Atr-Ⅱ) BMPR1A,BMPR1B	Saxophone,厚静脉 BMPR-Ⅱ	MAD Smad1(Smad5)
(OP-1)	BMP-5 Vgr-1/BMP-6 OP-1/BMP-7 OP-2/BMP-8	ActR-Ⅰ, BMPR-1A,BMPR-1B	Act-RII, Act-RIIB, BMPR-II,C14	
(GDF-5)	GDF-5/CDMP-1 GDF-6/CDMP-2 GDF-7			
(BMP-3)	BMP-3 GDF-10			
其他 BMPs	Vg-1/GDF-1 Dorsalin/Screw/BMP-9 Nodal GDF-3/Vgr-2			
未分类的	GDNF GDF-9 MIS		C14	

在不同的途径中有一定程度的交叉,可以看到,结合激活素受体可以是激活素和 BMPs。很多 TGF-β 的受体仍未被鉴定,还有些孤是儿受体。

缩写:TGF-β,转化生长因子-β;BMP,骨形态发生蛋白;Dpp,decapentaplegic(果蝇的 BMP-2 的同源蛋白);OP,成骨蛋白;CDMP,软骨衍生形态发生蛋白;GDF,生长/分化因子;GDNF,神经胶质衍生神经营养因子;MIS,Mullerian 抑制底物。

周期,参见*肿瘤抑制基因*),它们通过抑制细胞周期依赖蛋白激酶,阻止成视网膜细胞瘤易感蛋白的磷酸化,并阻断细胞周期进入 S 期。激活的Ⅰ型受体还在 C 端共同的 SSV/MS 基序上对 **SMAD 家族**蛋白磷酸化。然后 SMADs 形成异源二聚体,转移到细胞核内,与 DNA 结合蛋白作用或直接与 DNA 结合影响转录。在脊椎动物和无脊椎动物中发现了 10 多种 SMAD 蛋白,有些是传导途径特异的。Smad4 和果蝇中的同源蛋白 TGF-β 和 BMP 看来是信号途径中共同的。Smad6 和 Smad7 被认为是 SMAD 磷酸化的抑制蛋白,以负反馈起作用。

其他膜受体　　上面讨论的四种受体类型几乎代表了涉及动物细胞中信号传导的所有已知受体分子。有些信号是通过具有独特的信号机制受体传导。如心钠肽与有内在鸟苷酸环化酶活性的受体结合。这样增加了细胞内 cGMP 和活性蛋白激酶 G(PKG)的水平。许多有蛋白酪氨酸磷酸酶活性的受体被发现,包括白细胞共同的抗原 CD45 受体。肿瘤坏死因子受体家族用**死亡结构域**传导信号,死亡结构域与细胞质的类似结构的蛋白结合(参见*细胞凋亡*)。对其他从遗传水平分析的受体,它们的信号传导机制还远未阐明[如在神经系统中控制*侧抑制*(参阅)的 Notch 和 Delta]。

通过细胞质和细胞核受体的信号传导　大多数细胞外信号的受体是在细胞表面，因为信号分子是亲水性的，不能通过细胞膜。但一些可以直接穿过细胞膜的分子与细胞质或细胞核受体相互作用，并经常起始简单的信号传导途径。这种信号的例子包括脂溶性的类固醇、甲状腺激素和维甲酸以及气体氧化氮(参见受体鸟苷酸环化酶)。

类固醇激素受体超家族包括了类固醇、甲状腺激素和维生素 A 和 D 及其衍生物的受体，其中有重要的发育分子维甲酸。与配体的作用引起构象改变(**转化**)，刺激了 DNA 结合活性。有些受体位于细胞质中(如糖皮质激素受体)，转化使其易位到细胞核。其他受体已经位于细胞核内。被转化的受体是转录因子，通过高度保守的锌结合结构域与 DNA 相互作用(见第 17 章核酸结合蛋白)。它们可以根据识别位点的结构分成几个家族。有些受体的活性除配体结合外，还可被磷酸化调节(如孕酮和雌酮受体)，随之与其他信号途径汇合。

28.2　细胞内酶级联

细胞内信号　上面讨论了通过细胞膜的受体酪氨酸激酶或受体相关的酪氨酸激酶的活化，不需将配体转移到细胞中的信号传导机制。因为信号传导途径的最终目标经常不在细胞膜上，这样必须存在进一步的细胞内信号传导。这经常是通过依次激活的**酶级联**来介导：一个信号成分磷酸化并激活第二个成分，第二个成分再对第三个成分起同样作用，直到到达最终的目标。信号的过程也可以是抑制性的。

有些信号途径非常简单。JAK-STAT 和 TGF-β/SMAD 途径已经被阐述；都是涉及受体相关蛋白的磷酸化，然后进入细胞核，作为转录因子或它们的成分起作用。其他途径涉及很多步骤，提供了足够的分支和汇合的机会。不同途径的相互作用通过信号分子中重复基序来反映，使它们能与其他蛋白相互作用。有些基序在文框 28.1 中讨论。

Ras 途径　**Ras 家族蛋白**是膜结合的小蛋白，活性依赖于结合的鸟核苷酸辅助因子。Ras 本身是很多信号传导途径的中心成分。相关的 Rac 和 Rho 涉及对细胞运动和细胞骨架的控制。Ras 在活性(和 GTP 结合)和非活性(和 GDP 结合)状态间循环，这一过程由称为 **GAPs(GTP 酶活化蛋白)**和 **GNRPs(鸟核苷酸释放蛋白)**或 **GNEFs(鸟核苷酸交换因子)**的酶调节。GAPs 通过刺激 Ras 内在的 GTP 酶活性作用，然后使其失活，而 GNRPs 起相反作用。GAPs 和 GNRPs 都是通过 SH2 和 SH3 结构域与 RTK 信号复合物作用(图 28.2；文框 28.1)；这样信号传导可以刺激或抑制 Ras 活性，依赖于受体和每个活性状态酶的丰度。激活的 Ras 作用并刺激下游蛋白的活性，其中分析得最清楚的是丝氨酸/苏氨酸激酶 Raf(与 MAP 途径联系)。激活是通过将 Raf 带到细胞膜上(Raf 正常情况下是细胞质蛋白)，在细胞膜上 Raf 与可能是蛋白激酶的其他蛋白作用。持续结合在细胞膜上的重组 Raf 蛋白是持续激活的。这样将 Raf 带到细胞膜上对它的激酶活性是足够的。Ras 的其他靶分子被发现的极少；一个可能的候选者是磷酸肌醇 3-激酶(PI(3)K)(见第二信使，下文)。

在人类中发现了三个 Ras 基因(Ha-*ras*，Ki-*ras*，N-*ras*)，三个都是细胞或病毒癌基因(见第 18 章癌基因与癌)。

MAP 激酶信号级联　　MAP 激酶[促分裂原活化蛋白激酶，也称为 Erk(细胞外信号调节激酶)]是被很多生长因子激活的丝氨酸/苏氨酸蛋白激酶。MAP 激酶激活的途径包括 Ras 和 Raf 以及 MAP 激酶上游 Raf 磷酸化的激酶，称为 MAP 激酶的激酶(MAPKK)或 Mek(MAPK/Erk 激酶)(图 28.2)。MAP 激酶活化需要酪氨酸和苏氨酸的磷酸化，Mek 是**双特异性激酶**的一个例子，同时有酪氨酸和丝氨酸/苏氨酸激酶的活性。

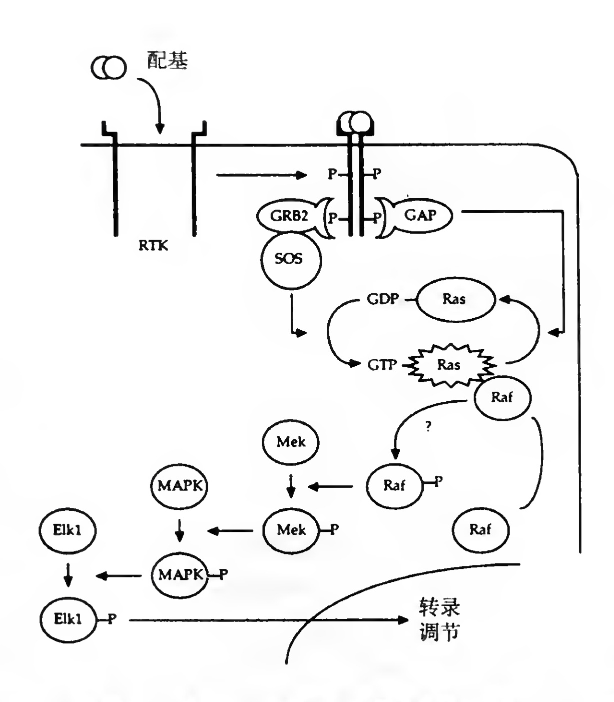


图 28.2　Ras-Raf-MAP 激酶级联。配体结合刺激受体酪氨酸激酶的二聚化和自身磷酸作用。有 SH2 结构域的蛋白识别磷酸化位点，并征集信号复合物。接头蛋白 GRB2 有 SH2 和 SH3 结构域；后者与鸟核苷酸交换因子(GNEF)SOS 结合。这样刺激了膜结合 Ras 的活化，Ras 征集 Raf。Raf 通过未知的膜相关激酶磷酸化，诱导了涉及 Mek 和 MAP 激酶的磷酸化级联。MAP 激酶磷酸化许多潜在的转录因子(图中以 Elk-1 为例)。不同的激酶活性被细胞质磷酸酶抑制，Ras 活性被 GTP 酶激活蛋白(GAP)抑制，GAP 同时被 RTK 信号征集，这样在信号传导后进行关闭。

MAP 激酶途径为有丝分裂信号(驱动细胞增殖的信号)开辟了从细胞表面进入细胞核的通道，途径的很多成分在不恰当活化(见癌基因与癌，细胞周期各章)时引发癌症。MAP 激酶的激活使其转移到细胞核中，并磷酸化和激活许多转录因子，包括 Elk-1、C/EPBβ 和 c-Myc(见下文)。MAP 激酶也磷酸化另一个称为 Rsk 的激酶，Rsk 转移到细胞核内，可以激活其他转录调节因子，如血清效应因子(SRF)。

MAP 激酶途径在真核生物中高度保守。与在哺乳动物中调节重要的生长反应一样，在果蝇和线虫中也发现了有同源成分的途径(它们控制细胞的分化；见发育：分子机制)，在酵母中的功能是对环境的监控(表 28.4)。在脊椎动物中，MAP 激酶途径的信号成分有相当的冗余，多个基因编码接头、Ras 激活和抑制蛋白、Ras 和 Raf、Mek 和 MAP 激酶本身。这些成分在底物专一性和细胞类型特异性或发育中的调节有差异，使它们在信号传导中起特异的作用。

表 28.4 RTK-Ras 途径通过 Raf 对 Mek 的磷酸化与 MAP 激酶途径联系

种类	酿酒酵母	栗酒酵母	线虫	黑腹果蝇	哺乳动物
受体			Let-23	Sevenless	很多
接头			Sem-3	Drk	Grb2
GNRF				Sos	SOS
Ras			Let-60	Ras1	Ha-Ras, Ki-Ras, N-Ras
Raf			Let-45	D-Raf	c-Raf, A-Raf, B-Raf
Mek	STE7	Byr1	Mek-2	D-Sor	Mek1, Mek2
MAPK	FUS3, KSS1	Spk1	Mpk-1/Sur-1	Rolled	Erk1, Erk2
靶分子	STE12, FAR1		Lin-1, Lin-31	Pointed Rsk	c-Fos

酵母缺乏 RTK，但信号由 G 蛋白偶联受体通过 Mek 激酶激活 MAP 激酶途径传导。酵母途径是传导交配信息素信号，线虫途径涉及阴门专一性，果蝇途径涉及眼发育过程中 R7 光受体细胞专一性。哺乳动物途径有很多成分冗余。

应激激活激酶　　**应激激活蛋白激酶（SAP 激酶）**是与 MAP 激酶相关的丝氨酸/苏氨酸激酶，它在如紫外线照射和感染细胞因子的应激反应中激活转录因子 c-Jun，但对生长刺激反应很小。SAP 激酶途径在 T 细胞发育中也被激活，并在有些细胞中可以诱导生长。已知至少有八种 SAP 激酶蛋白，通过不同的剪接从三个基因中得到。单个的 SAP 激酶命名为 Jnk-1、Jnk-2 等，反应了它们的底物专一性：它们磷酸化 Jun 的 N 端 Ser63 和 Ser73 残基（因此是 Jun 的 N 端激酶），以及如 Elk-1 和 ATF-2 的其他转录因子。

像 MAP 激酶一样，SAP 激酶激活需要酪氨酸和苏氨酸的磷酸化。推断的 MAP 激酶相关分子的几个家族还未被发现，它们在双磷酸化位点上有差别。如在 MAP 激酶中，被识别序列是 Thr-Pro-Tyr 序列，而 SAP 激酶是 Thr-Glu-Tyr 序列。位于 SAP 激酶上游的双特异性激酶称为 Sek（SAPK/Erk 激酶），与酵母蛋白 STE7（酿酒酵母）和 Byr1（栗酒酵母）紧密相关。根据与 STE7 和 Byr1 上游激活因子 STE11 和 Byr2 同源性从哺乳动物中分离的蛋白开始被称为 MAP 激酶的激酶的激酶（或 Mek 激酶），因为它在过量表达时，在体外和体内能够磷酸化 Mek。现在认为 Mek 激酶在生理条件下并不磷酸化 Mek，而是对 Sek-SAPK 途径特异的。哺乳动物 Mek 激酶的上游蛋白正在被研究。

调节信号信息的细胞质激酶和磷酸酶网络　　正如上文讨论的，对信号的反应不仅依赖于信号本身（反映了不同信号成分的有效性）和反应细胞的状态，而且依赖于信号的耐受性。这反应了激活后关闭信号所需的时间。信号传导途径的减弱是关键的，这样信号机制才不会饱和，对于信号途径的每一个活性成分，都有一个抑制因子。抑制因子可以被认为以几种方式起作用。

（1）当信号传导途径被激活，其中的一个激活成分经常是相同途径的抑制因子。抑制因子比激活因子作用晚一些，这样信号不会在信息转移之前被关闭。

（2）最初信号的发起依赖于途径中激活因子和抑制因子之间的平衡。

（3）有些信号的目的是通过激活特定的抑制因子，抑制其他信号的传导。

细胞可以被认为是一个加工中心，到达它表面的信息转换成有相反活性的不同分子的合成和激活，如激酶和磷酸酶（图 28.3）。信号途径被交连和反馈环路的复杂系统所控制，这种系统依赖于细胞内不同力量之间的平衡。新的信息到达打破了细胞中维持的平

衡,引起如特定激酶瞬时的活性比相反的磷酸酶活性更强。这样在信号通过反馈和新的平衡建立而关闭前,使激酶活性爆发,在给定的转录因子或酶的激活时活性达到最高值。

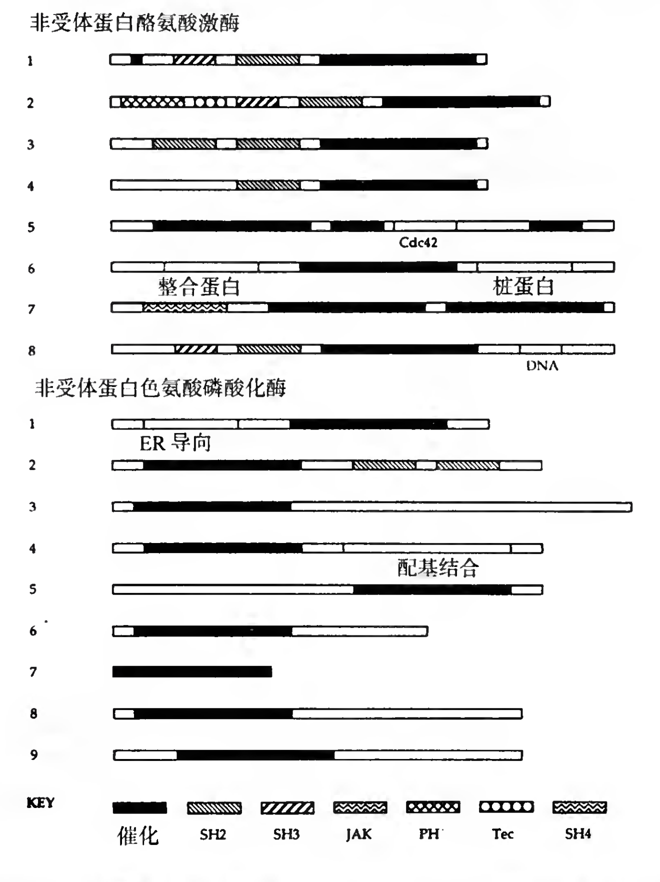


图 28.3 细胞质蛋白酪氨酸激酶和磷酸酶,负责控制细胞内信号的复杂的调节途径。激酶根据结构域结构和大小分成家族。“1”代表了大小不同的四个家族的(Csk,Sm,Brk/Rak 和 Lyn/Lyk/Blk)的共同功能域结构;只有 Lyn/Lyk/Blk 家族有 SH4 结构域。2,Tec 家族(Tec,Src,Yes,Fyn);3,Syk/Zap 家族;4,Fes 家族;5,Ack;6,Fak;7,Janus 家族;8,Abl/Arg。磷酸酶根据它们非催化结构域分类:1,PTP1B;2,SHPTP;3,PTP1H;4,MEG2;5,PTP-PEST;6,YOP 2b;6,VH1;7,Cdc 25A;8,MKP-1。6~8 是双专一性磷酸酶;MKP-1 被认为是 MAP 激酶的磷酸酶,灭活 MAP 激酶。特殊的结合结构域在合适的位置表示。

28.3 第二信使

第二信使的概念 细胞对广泛的信号反应需要大量的受体。然而,反应的范围要小得多。很多信号到达细胞表面,如引起细胞分裂或从细胞周期中退出。其他信号诱导特定基因的表达,保护细胞不受应激的影响。因此,早期的信号途径交汇成细胞内少量的信号网络,使细胞将到达细胞表面的复杂信息转换成细胞质中简单的生化信号。涉及这一过程的分子称为**第二信使**。

环核苷酸第二信使 最早发现的第二信使是环核苷酸 cAMP 和 cGMP。这些分子在细胞中的水平由核苷酸环化酶的相反活性控制,核苷酸环化酶催化 NMP→cNMP 的反应,

而环核苷酸磷酸脂酶(PDEs)催化反向反应。几种早期信号途径影响核苷酸环化酶的活性。G蛋白偶联受体的 G_s 和 G_i 家族刺激和抑制腺苷酸环化酶,相反传导素(Gt)刺激 cGMP PDE 活性。少数细胞表面受体有内在的鸟苷酸环化酶活性。氧化氮经亚铁血红素结合后也有受体鸟苷酸环化酶活性。

环核苷酸下游是活性依赖于环核苷酸的效应分子:依赖于 cAMP 的蛋白激酶 A(PKA)和依赖于 cGMP 的蛋白激酶 G(PKG),它们是具有多种不同底物的丝氨酸/苏氨酸激酶(见下文)。也有环核苷酸门控(cyclic nucleotide-gated)的离子通道和涉及环核苷酸依赖的 PDEs 的反馈途径。其他 PDEs 通过磷酸化或钙调蛋白(见下文的钙信号)被调节。细胞中对环核苷酸水平的控制在图 28.4 中总结。

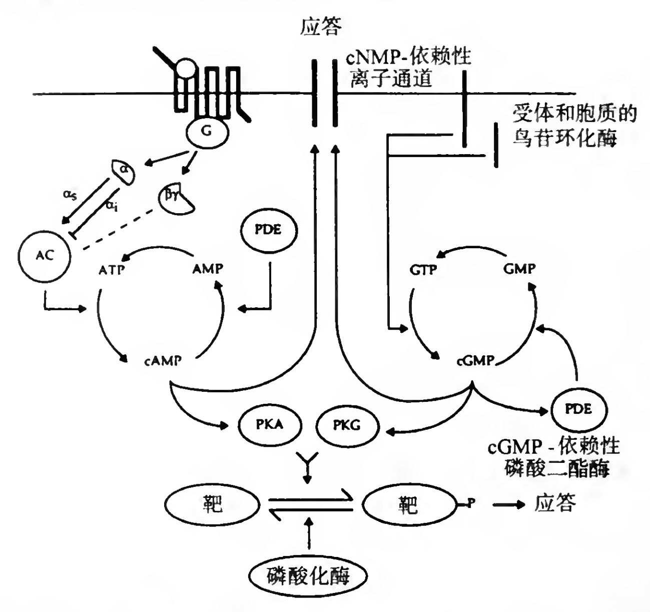


图 28.4 环核苷酸第二信使。腺苷酸环化酶合成 cAMP 受到 G 蛋白 α_s 亚基刺激而被 G 蛋白 α_i 亚基抑制(βγ 亚基也可能影响 AC 活性);cAMP 被 cAMP 磷酸二酯酶从细胞中去除。鸟苷酸环化酶跨膜结构域或细胞质受体合成 cGMP。cAMP 和 cGMP 都激活环核苷酸依赖的激酶和离子通道。cGMP 还激活在负反馈中一 cGMP-依赖的 PDE。

蛋白激酶 A 在细胞中磷酸化很广范围的蛋白,包括很多酶、受体、离子通道和转录因子。后者的一个例子是转录因子 CREB,CREB 与称为 cAMP 效应元件的顺式作用元件结合(见下文)。少数的 PKG 的靶分子被确定,包括几种 G 蛋白和 Ins(1,4,5)P_3 受体。PKA 和 PKG 以同源二聚体发挥作用,PKA 与催化和调节亚基形成的四聚体是非活性状态的,后者在 cAMP 结合后释放。PKG 的催化和调节亚基是单链多肽的一部分。PKA 和 PKG 有几种同工酶,显示出不同的底物特异性。另外两种酶之间有一定的交叉,如在特定细胞中 PKA 可以被 cGMP 激活,而 PKG 被 cAMP 激活。PKA 和 PKG 活性被一系列不同底物特异性的蛋白磷酸酶和复杂的转录和转录后调节途径所反转。细胞对特定信号的反应依赖于特定激酶和磷酸酶活性的平衡。

作为第二信使的脂类　得到深入研究的第二信使体系包括细胞膜内磷脂水解的产物,**磷脂酰肌醇-4,5-二磷酸[PtdIns(4,5)P_2]**。G 蛋白偶联受体相联系的 G_q 家族蛋白(G_q,

G_{11},G_{12})的激活刺激磷酸酶 C-β(PLC-β)的活性。PLC 酶将 PtdIns(4,5)P_2 剪切成两个成分,**肌醇-1,4,5-三磷酸(Ins(1,4,5)P_3)**和**1,2 二酰甘油(DAG)**。另一个 PLC 同工酶,PLC-γ 通过它的 SH2 结构域转移到细胞膜,被 RTK 信号激活。Ins(1,4,5)P_3 和 DAG 都是重要的第二信使(图 28.5)。

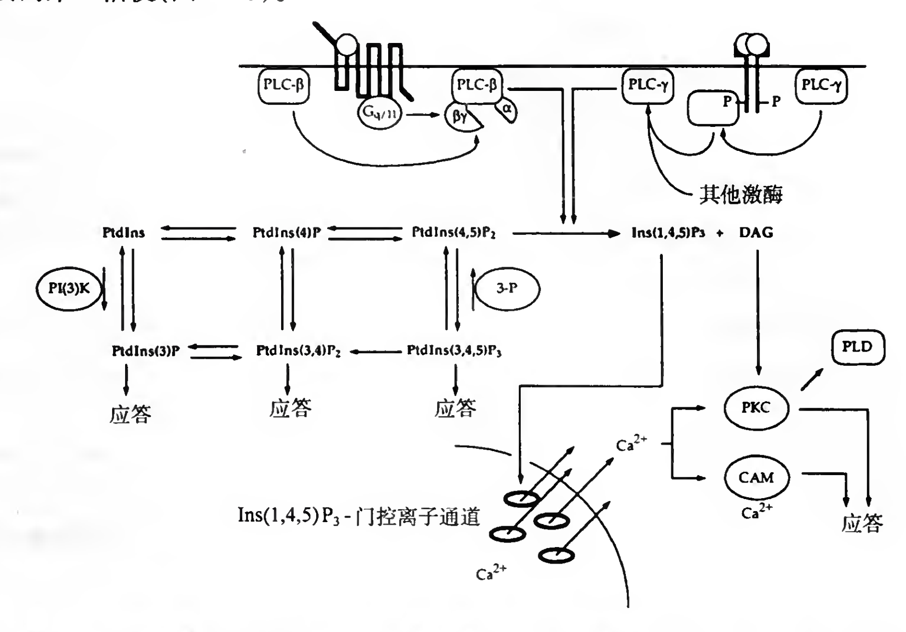


图 28.5 细胞中磷脂和钙离子第二信使。磷脂酶 C(PLC)将脂膜成分 PtdIns(4,5)P_2 转换成肌醇-1,4,5-三磷酸[Ins(1,4,5)P_3]和 1,2 二酰甘油(DAG)。不同形式的 PLC 被 Gq 蛋白和激活受体和非受体激酶的磷酸化残基刺激。Ins(1,4,5)P_3 激活内质网(ER)膜上 Ins(1,4,5)P_3 依赖的钙离子通道释放钙离子进入细胞质,与此同时 DAG 激活磷酸激酶 C(PKC)。PKC 有很多底物,包括磷脂酶 D(PLD)。Ca^{2+} 与钙调蛋白(CaM)结合,激活 Ca^{2+}/CaM 依赖激酶。磷脂酰肌醇(PtdIns)能够在几个位置上磷酸化,并且五种磷酸化衍生物可以相互转换。磷酸肌醇 3-激酶[PI(3)K]在 3D 位置上加上磷酸基团,起始一系列下游反应。这个过程被 3-磷酸酶(3-P)逆转。

Ins(1,4,5)P_3 通过激活 Ins(1,4,5)P_3 门控的钙离子通道,介导钙离子从内质网或肌质网中释放(参见钙离子信号,下文)。这样的效果可以被**离子载体**模拟:离子载体如 A23187,像持续开放的钙离子通道一样起作用。有些 Ins(1,4,5)P_3 可以进一步磷酸化为肌醇-1,3,4,5-四磷酸[Ins(1,3,4,5)P_4],其效应仍是未知的。Ins(1,4,5)P_3 还可以与不同蛋白特异性作用,包括磷脂酶 C-δ 和 SOS,起始不同的信号级联。DAG 与钙以及其他膜磷脂磷脂酰丝氨酸形成化合物激活一系列钙依赖蛋白激酶(PKC)的同工酶。这种效果可以被植物衍生的直接激活 PKC 的佛波酯(phorbol esters)所模拟。至少有 11 种 PKC 的同工酶,根据结构域结构分成四种主要家族。它们在细胞中是根据细胞类型特异性和分化分布的。基因敲除、过量表达和选择性抑制的研究表明它们有重叠的底物特异性,在不同的细胞中有不同的调节效果,但这样的研究破坏了体内亚细胞定位,对确定生理状态下每种酶的特异性目标分子有困难。PKC 的一个底物是磷脂酶D(PLD),PLD 通过末端磷酸二脂键水解膜磷脂,产生磷脂酸(PLD 也可以被 G 蛋白激活)。

另一个细胞内脂类信号体系涉及磷脂酰肌醇在 D3 位置的磷酸化。**磷酸肌醇 3-激酶[PI(3)K]**在碱性的 PtdIns 和 PtdIns(4)P、PtdIns(4,5)P_2 加上 D3 磷酸基团产生 PtdIns

(3)P、PtdIns(3,4)P_2 和 PtdIns(3,4,5)P_3(图 28.5);有三类 PI(3)K 使用不同磷脂酰肌醇作为底物。这些脂类在未刺激的细胞中很少,但在细胞外信号和肿瘤转化后积累。它们特异地或以不同亲和力与很多下游效应分子结合,将其转移到膜,起始第二信号级联反应。PtdIns(3,4,5)P_3 可能是这些第二信使中最重要的:它通过 SH_2 和 PH 结构域与多个靶分子作用(如 Src、PI(3)K,磷酸激酶 C-ε 和 λ 同种型(λ isoforms),以及特异性的 PtdIns(3,4,5)P_3 依赖蛋白激酶 PDK-1 和 PDK-2)。PtdIns(3,4)P_2 靶分子与 PtdIns(3,4,5)P_3 重叠,但 PtdIns(3)P 表现出与蛋白运输有关的衔接蛋白的特异性结合。

脂类第二信使也可以在细胞间传递信号。磷脂酶 A2 合成花生四烯酸和溶血磷脂,它们是细胞内第二信使。然而它们也调节了前列腺素和白三烯的产生。有九种不同类型的前列腺素,由很多不同的细胞类型合成和分泌。它们以旁信号起作用,而不像大多数脂类信号蛋白,与细胞表面的受体结合。

作为第二信使的钙离子 细胞内钙离子水平由不同的过程控制,包括代谢过程、生长和像膜兴奋和肌肉收缩等特殊功能。钙离子信号途径涉及被钙离子激活的钙离子结合蛋白,它与非活性形式蛋白相结合,并刺激它们而影响许多下游途径。细胞质中 Ca^{2+} 水平通过外运保持在较低的水平,但在细胞外和特定的细胞内间隔中有其 10000 倍的浓度。从细胞表面起始的上游信号通过开放钙离子通道诱导 Ca^{2+} 释放进入细胞质中,这样增加激活的钙离子结合蛋白的量,诱导下游信号途径(表 28.5)。刺激钙离子释放的两个主要途径是 G 蛋白偶联受体相关 G_q 家族蛋白和 RTK。这些受体都刺激增加 Ins(1,4,5)P_3 的磷脂酶(图 28.5),导致 ER 膜上 Ins(1,4,5)P_3 门 Ca^{2+} 通道的开放。

表 28.5 细胞质中钙离子水平的调节

Ca^{2+} 运输	调节因子
从细胞外空间运入	在膜去极化后,电位门 Ca^{2+} 通道释放 Ca^{2+} 进入细胞 在配体结合后(如 NMDA 受体),配体门 Ca^{2+} 通道释放 Ca^{2+} 进入细胞 CRAC 通道(钙离子释放激活 Ca^{2+} 通道被从 ER 中释放的 Ca^{2+} 激活)
从 ER/SR 运入	Ins(1,4,5)P_3 门通道 Ryanodine 受体通过 NAD^+ 衍生环腺苷二磷酸核糖(cADPR)的激活释放 Ca^{2+} Ca^{2+} 通过激活上述通道释放 Ca^{2+}
运出	Na^+ / Ca^{2+} 交换——用顺电化学梯度运输 Na^+ 的能量逆电化学梯度外运 Ca^{2+} 细胞质和 ER/SR 上的 Ca^{2+}ATP 酶从细胞质中外运 Ca^{2+}

细胞质中有很多不同的钙离子结合蛋白。尽管大多数起缓冲作用(减低细胞质中 Ca^{2+} 浓度的另一机制),有些是作为钙离子调节和信号蛋白作用。**膜联蛋白**(annexins)[1)]是钙离子依赖膜结合蛋白,可以改组细胞骨架的因子;它们也抑制磷脂酶 A2。主要的调节钙离子结合蛋白属于 EF-hand 超家族(见蛋白质:结构、功能和进化),以及**钙调蛋白**(calmodulin, **CaM**)和肌钙蛋白 C。CaM 是普遍存在和多功能的蛋白,肌钙蛋白 C 是肌肉特异性的,在肌肉收缩时控制肌动蛋白和肌球蛋白的相互作用。CaM 激活许多蛋白激酶(CaM 激酶Ⅱ、延伸因子激酶),磷酸酶(钙调磷酸酶)和细胞骨架因子,并且还激活 Ca^{2+}

1)一类被钙离子活化后可与膜磷脂结合的蛋白,参与膜转运及膜表面其他一系列依赖于钙调蛋白的活动,分为Ⅰ,Ⅱ,Ⅲ…等多种。其中膜联蛋白[即脂皮质蛋白](lipocortin Ⅰ),膜联蛋白Ⅱ即依钙结合蛋白(calpactin),膜联蛋白Ⅳ即内联蛋白(endonexin),膜联蛋白Ⅴ即内联蛋白Ⅱ或锚定蛋白(anchorin),膜联蛋白Ⅵ即钙磷脂结合蛋白(calphobindin)或钙电蛋白(calelectrin),膜联蛋白Ⅶ即会联蛋白(synexin),膜联蛋白Ⅺ即会联蛋白Ⅱ。——译者注

ATP 酶，促进 Ca^{2+} 从细胞质中移出。CaM 也与其他第二信使系统的成分作用，包括腺苷环化酶、IP(3)K和氧化氮合成酶。

串话　几个得到细致研究的信号途径和第二信使系统在以上章节中已被阐述，但正如所讨论的，途径之间有广泛的相互作用，因此一个刺激可以激活细胞中几个途径，而不同的信号可以产生相同的效果。这可以几种机制发生。

(1)受体-配体混杂，配体激活多个受体或多个配体激活相同的受体，或者受体包含多种不同信号特异性的多聚体。

(2)分歧，受体的刺激激活两个平行的途径。一个例子 PTK 活性对 Ras-Raf-MAP 激酶途径和磷脂酶 C-γ[增加第二信使 Ins(1,4,5)P_3 和 DAG 水平]的激活。这通过不同信号分子所拥有的结构域调节，在所举例子中是 SH2 结构域与激活受体作用。

(3)**串话**(cross-talk)，是一个途径的分支与另一个途径作用(图 28.6)。细胞中所有主要的信号途径都使用蛋白激酶和磷酸酶。它们之间有大量的相互作用，如蛋白激酶 A 被 G 蛋白激活，调节增加 cAMP 水平，灭活 Ras。相反蛋白激酶 C 是钙离子依赖的，刺激 Ras。细胞中主要的第二信使系统也是相互依赖的：Ins(1,4,5)P_3 诱导钙离子从 ER 中转运，DAG 与钙离子一起激活蛋白激酶 C。钙离子依赖分子 CaM 调节 PI(3)激酶和腺苷环化酶活性，它们分别再控制 Ins(1,4,5)P_3 和 cAMP 水平。cAMP 可以激活离子通道，因此影响细胞质中 Ca^{2+} 水平。

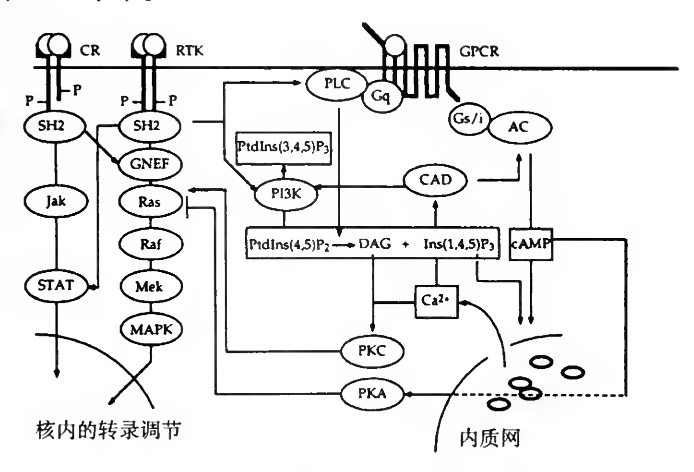


图 28.6　信号途径中串话的举例。MAP 激酶途径可以被 RTK 信号通过 Ras 和 Raf 激活，但 Ras 也可以通过由钙离子调节的细胞因子受体和蛋白激酶 C 的串话被激活。有些 RTK 也可以直接调节 STAT(如 EGF 受体)。Ras 被蛋白激酶 A(PKA)灭活，PAK 被 cAMP 激活。细胞中主要的第二信使系统也是相互依赖的，Ins(1,4,5)P_3 和 cAMP 都影响钙离子释放，钙离子通过钙离子依赖分子 CaM 控制磷脂酰肌醇 3 激酶(PI(3)K)和腺苷环化酶(AC)的活性。细胞质中包含很多其他激酶和磷酸酶，将不同的信号途径连成一复杂的网络。蛋白质用圆形表示，第二信使以矩形表示。CR，细胞因子受体；RTK，受体酪氨酸激酶；GPCR，G 蛋白偶联受体。

这些无数的相互联系添加在广泛的激酶和磷酸酶的调节网络上，让人怀疑任何信号特异性是否得到维持；信号特异性的机制是目前研究的主要课题。然而，细胞只合成所讨论的很多信号分子中一部分，所以单个细胞可以只对某些信号反应，并通过细胞中特定成分和活性限制途径进行反应。另一个可用于调节信号反应机制是广泛的基因表达调节。这样细胞核中激活的转录因子通过选择性外遗传沉默将获得不同的基因组合(见染色质，

DNA 甲基化及外遗传调节各章)。

28.4 信号交付

对信号的反应 到达细胞表面的不同的信号转化成有序的激酶活性的级联反应,或产生第二信使,第二信使本身通过调节细胞激酶的活性发挥它们的效应。激酶的级联反应被磷酸酶控制,磷酸酶的丰度和活性由复杂的调节所控制。信号交付(signal delivery)可定义为信号传导途径的末端,即起始信号的细胞靶成分被磷酸化。在这方面,磷酸基团是广泛的引起蛋白形状和活性改变的信号流通物,尽管同样的效果可以通过其他形式的修饰而产生(参见*组蛋白乙酰化*,*蛋白质修饰*)。信号以两种方式调节它们的最终效果。

(1)通过调节细胞中已经存在蛋白的活性。如蛋白激酶 A 磷酸化几个在代谢中关键的酶(如肌细胞中它磷酸化(和抑制)糖原合成酶,但导致糖原磷酸酶的激活,使细胞使用糖原)细胞骨架蛋白也受钙调蛋白依赖激酶的调节(如神经调节蛋白,Tau;另参见 M *期激酶*)。

(2)通过调节转录调节因子或翻译调节因子的活性,从而影响基因表达。

这些效应可以通过用转录或蛋白合成的抑制剂,阻断从头进行的基因表达加以区分。

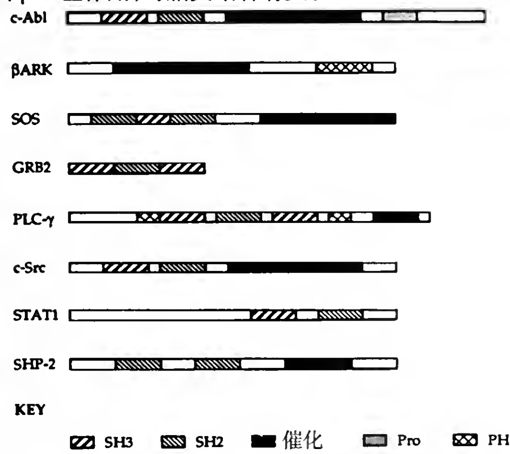
转录因子的激活 很多信号传导途径在细胞核中终止。下游的靶可以是细胞周期机制的成分,控制细胞生长或转录因子,转录因子通过靶基因的*效应元件*(参阅)引出基因表达的特殊模式。很多转录因子可以被特殊的信号途径激活,在很多情况下,反应的多样性通过转录因子成分不同的同工形式增加。一些研究得较清楚的途径在表 28.6 中显示。在受体刺激的六小时中,大约 100 个早期基因被激活。这些包括某些原癌基因(如 c-*jun*,c-*fos*),很多其他转录调节因子的基因(如 *ets*、*srf*、激素受体)和结构蛋白(纤连蛋白、肌动蛋白)。反应中激活的很多基因是生长促进信号(如 RTK-Ras-Raf-MAP 激酶)和生长抑制信号(如 TGF-β-CDI)。

表 28.6 信号交付到转录因子

转录因子	信号交付机制(激活)
AP1(c-Jun)	被 MAP 激酶激活的 Rsk-1 抑制特定激酶,将 DNA 结合结构域去磷酸化 通过 SAP 激酶在反式激活结构域上磷酸化 Ser-63 和 Ser73
(c-Fos)	通过 Elk-1 激活诱导 c-*fos* 基因表达(见*分离通路*) FRK,一种 Ras 激活的激酶不同于 MAP 激酶,对反式激活结构域中的 Thr-232 磷酸化 (C 端丝氨酸/苏氨酸的磷酸化引起 c-Fos 关闭 c-*fos* 基因的表达)
CREB	PKA 在 Ser-133 的磷酸化使其与起始复合物作用的辅助蛋白 CREB 结合蛋白(CBP)作用,
CREM	反式激活需要 Ser-117 的磷酸化
Dorsal(NF-κB)	丸状激酶的磷酸化促进抑制因子 Cactus(I-κB)的释放,显露出核定位顺序,并转移到细胞核中
Elk-1/SRF	MAP 激酶对 Elk-1 C 端丝氨酸残基的磷酸化(a),使其能够与 SRF 和其他调节成分形成稳定的复合物,并(b)刺激反式激活结构域
p53	DNA 损伤引起 p53 mRNA 增加翻译并增加蛋白质的稳定性
STATs	磷酸化促进二聚化,并转移到细胞核中
类固醇受体超家族	配体结合引起构象改变,允许与 DNA 结合 有些受体也可以被丝氨酸残基的磷酸化调节

文框 28.1 Src 同源性和其他保守的信号结构域

信号蛋白的结构域结构 很多信号传导蛋白包含了保守的结构域，有利于不同信号传导途径成分间的相互作用。这在非受体酪氨酸激酶的 Src 家族中作为共同元件最先发现，称为 **Src-同源(SH)结构域**。SH1 本身是激酶结构域，是最保守的元件。SH2 在很多酪氨酸激酶和其他分子中发现(见图)——它是磷酸酪氨酸结合基序，也就是说它允许蛋白与磷酸化的酪氨酸结合，如在激活的 RTKs 中的磷酸酪氨酸残基。SH3 是进一步的结合结构域，倾向于和脯氨酸富集的基序作用。SH4 是调节甘氨酸的豆蔻酰化的短结构域，因此允许膜的对接(只有 Src 家族激酶和 AbI 的特定同种型有 SH4 结构域)。其他几个保守的信号元件也被发现了，包括与 SH2 不同的磷酸酪氨酸结合结构域(PTB)和血小板-白细胞激酶 C 底物同源结构域(PH)，PH 的确切功能还不清楚，但它通过 WD40 基序与 G 蛋白 β 亚基作用，与脂类结合有关。



图显示了有保守结构域 SH2、SH3、Pro(脯氨酸富集，假定的 SH3 识别结构域)和 PH(血小板-白细胞激酶 C 底物同源结构域)的信号蛋白的差异。c-AbI 和 c-Src 是激酶，PLC-γ 是磷脂酶，SHP-2 是磷酸酶，β ARK 是受体酪氨酸激酶，STAT1 是转录因子，SOS 是鸟苷酸交换因子，GRB2 是接头蛋白。注：PLC-γ 的 PH 结构域被 SH3-SH3-SH3 组件打断。

(陈 淳 译)

第29章
转　　录

基本概念和定义

- **转录**是以 DNA 为模板合成 RNA。在体内,转录是基因表达的第一阶段并且是基因调节的主要阶段。另外 DNA 复制过程中的 RNA 引物也是通过转录产生的。转录在反转录病毒(retroid viruses,参见)复制周期中起重要作用。几种细菌和真核生物转录系统中的成分已经被发现,这样允许对已克隆的基因在体外进行转录翻译偶联。
- 与 DNA 复制不同,转录是不对称的过程——只有 DNA 的一条链作为模板(图 29.1)。新生 RNA 与 DNA 复制中的前导链(参见)相类似,是连续转录的。每个 RNA 分子都是**转录本**(transcript),DNA 上的转录区域称为**转录单位**(transcription unit,参见基因、顺反子、操纵子)。转录的第一个核苷酸定义为转录单位的 +1 位。对应 DNA 链的紧接着上游的核苷酸为 −1 位,没有 0 位。
- 催化转录的酶称为(**DNA 依赖的)RNA 聚合酶**。与 DNA 聚合酶不同,RNA 聚合酶从头起始链的合成(没有引物)且不校正转录本(参见文框 26.1)。
- 成功的转录需要 RNA 聚合酶作用于激活基因上游的顺式作用启动子位置。原核生物 RNA 聚合酶直接与 DNA 结合而真核生物需要其他蛋白在启动子处形成起始复合物。RNA 聚合酶的结合效率受转录因子正或负的作用,还受转录因子与启动子结合或与更远的增强子(正)或沉默子(负)位点结合的影响。增强子/沉默子位点通过中间区域 DNA 的突出成环与起始复合物相互作用。

29.1　转录的基本原理

转录的阶段　同其他多聚化反应一样,转录分为三个阶段:**起始**,RNA 聚合酶与 DNA 结合并开始合成 RNA;延伸,RNA 链被延伸,RNA 合成主要发生在此阶段;**终止**,延伸终止,转录本从模板上脱离的阶段。原核和真核生物的具体机制有所不同,但基本原理是类似的。

在起始阶段,RNA 聚合酶识别基因起始位点附近的顺式作用元件,**启动子**,并与其结合,起始的结合产生一**封闭的启动子复合物**(closed promoter complex)。成功的起始将其转变成**开放的启动子复合物**(open promoter complex),在这一复合物中 DNA 部分解链。转录 以插入第一个核苷酸(经常是嘌呤核苷酸)开始。起始的终止以**启动子的清理**(promoter clearance)为标志,RNA 聚合酶离开启动子位点但不解离,留出启动子以进行下一

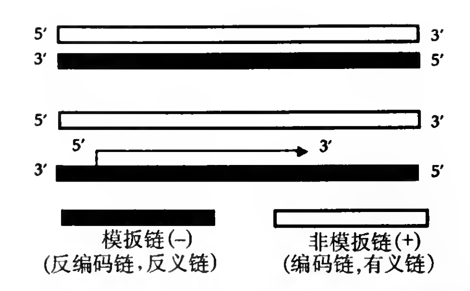


图 29.1 转录单位中 DNA 链的名称。因为只有一条链作为转录的模板,所以需要区分 DNA 的两条链。**模板链**(template strand)是与转录本互补,而**非模板链**(nontemplate strand) 除了以胸腺嘧啶代替了尿嘧啶与转录本携带相同的核苷酸。在编码蛋白的基因中,非模板携带与 mRNA 相同的信息,称为**编码链**(coding strand)、**有义链**(sense strand)或(+)链,而模板链是**反编码链**(anti-coding strand)、**反义链**(antisense strand)或(-)链;反义链与反义 RNA 有相同的序列。基因组 DNA 因邻近基因方向不同,有义链是可以相互转换的。只有在特定的病毒基因组中所有的基因都是一个方向的,这样整个基因组可以被定义为(+)或(-)链(参见*双义 RNA*、*无义密码子*、*无义突变*和*反转录 RNA*)。对有义和无义定义 DNA 链还有不同的看法。这儿的定义是从逻辑上而言的(有义 DNA 链 = mRNA 序列,反义 DNA 链 = 反义 RNA 序列),与其最初的用法有所不同。

次转录起始。启动子的清理只有在开放的启动子复合物是稳定的情况下才发生,且经常伴随一些产生短转录本的**流产起始**(abortive initiations)。这是 RNA 聚合酶的普遍属性,是从头链合成所要求的。起始是转录过程中的限速步骤,在原核和真核生物都是基因调节中首要阶段。调节因子与启动子附近的顺式作用位点或更远处的位点结合,与起始蛋白作用,以调节酶的起始结合、稳定开放启动子复合物的形成或启动子清理的效率。

一旦转录成功起始后,RNA 聚合酶就沿着模板以 5′→3′方向合成新生 RNA。双链 DNA 在延伸复合物前解链并在其后重新形成双链结构;在真核生物中,还涉及 DNA 核小体结构(参见*染色质*)的破坏。转录本与模板的配对是短暂的。转录位点处的动态、短暂的解链结构称为**转录泡**(transcription bubble)。延伸的速度是恒定的,只是偶尔可以被模板的二级结构所打断(参见*共转录调节*、*转录协同修复*)。在原核生物中,持续的延伸依赖于同步的蛋白合成(参见*衰减*,attenuation)。

转录终止有几种可能的机制:新生转录本上的二级结构、DNA 上终止位点或转录本的剪切。转录的终止涉及酶和所有相关因子的释放和新生转录本的解离。

转录活性的三种成分 **转录活性**(transcriptional activity)是 RNA 合成的速度,即每分钟合成转录本的长度。转录起始经常是限速步骤,所以转录活性直接反映了转录起始的效率。转录起始的调控分为三种成分:基本的、结构性和可调节的。

基本成分(basal components)初步的作用是帮助 RNA 聚合酶定位于转录起始点处,使 RNA 合成可以起始。

结构性成分(constitutive components)控制在没有其他调控情况下的起始效率,决定了每个基因的基本转录活性,允许不同的基因以不同的速率表达以满足细胞的基本需要。

调节成分(regulatory components)改变起始效率,使个别基因对环境改变作出反应而产生不同的转录活性。在多细胞生物中,转录调节在细胞分化的特化和维持中也起了重要作用。

如下文将讨论的,原核和真核生物中这三种转录调节成分不同(见表 29.1)。

表 29.1 真核生物与原核生物转录起始调控的比较

	细菌	真核生物
基本启动子元件	-10 和 -35 序列(起始位点上下游也对起始有作用)	Ⅰ型:核心启动子 Ⅱ型和一些Ⅲ型:不常见的启动因子或 TATA 盒 Ⅲ型:内部由二部分构成的启动子
识别	RNA 聚合酶的 σ 因子	TBP 和其他 GTFs
组成型表达的控制	依赖于整个基本启动子序列、上下游序列和结构	依赖于基本启动子结构和结合于启动子/增强子的顺式作用位点的基本转录活化子
调控表达	主要由结合在启动子附近或重叠位置调节元件的蛋白调节。远距离的增强子很罕见	Ⅰ型基因表达是恒定的。Ⅱ型和Ⅲ型基因受启动子上游调节元件和远处增强子/沉默子元件的结合蛋白调节,这些调节元件通过使内部 DNA 成环突出而与启动子作用
策略	功能相关基因经常以操纵子形式成簇排列和共转录。基因排列紧密允许共同转录调节。快速反应的策略是很重要的,因此转录因子调节中的变构调节、反义 RNA 调节、不稳定 mRNA 和转录与蛋白合成的关联调节	功能相关基因经常为重叠的调控元件所分散(与 β-珠蛋白家族、*Hox* 基因比较)-没有操纵子结构(参见文中的例外情况),但也有一些协同的顺式调节的例子(例如并参见基因座控制区域)。基因高度分散-共转录情况很少(与亲本印记比较)。快速和缓慢反应都很普遍。快速反应往往与信号传导有关,因此主要是转录因子的共价修饰。缓慢反应包括分化,经常涉及转录因子的重新合成

TBP:TATA-结合蛋白;GTF:一般转录因子(general transcription factor)。

29.2 原核生物的转录起始——基本的结构成分

RNA 聚合酶 在原核生物中,一个 RNA 聚合酶转录所有的基因。大肠杆菌 RNA 聚合酶是四聚体的**核心酶**(core enzyme),包含 α 和 β 亚基,分子式为 $\alpha_2\beta\beta'$。这对转录延伸已足够,但起始还需要一个称为 σ 的亚基,σ 亚基加入组成**全酶**(holoenzyme)。**σ 因子**有两个功能:它识别启动子并将封闭的启动子复合物转换成开放的状态。一旦转录起始,σ 因子从全酶中脱离。核心酶能与 DNA 结合,但效率低,特异性差。σ 因子的基本功能是增加 RNA 聚合酶对启动子的特异性结合,减少非特异性结合。

一个 σ 因子(大肠杆菌中是 σ^{70})起始大部分基因的转录,而其他 σ 因子识别不同的启动子,可以诱导特殊基因的协同表达(如涉及热休克反应的基因)。一些噬菌体(如 T4)编码自身的 σ 因子,这种 σ 因子使宿主细胞的核心酶转录噬菌体的基因。其他噬菌体(如 T3、T7)则编码自己的 RNA 聚合酶,是对噬菌体启动子有强亲和力的单链多肽。

其他细菌与大肠杆菌有类似的 RNA 聚合酶,尽管亚基数目可能有所不同。相反,古细菌的 RNA 聚合酶与真核生物类似。

基本转录起始 细菌中转录起始是 RNA 聚合酶的全酶与位于基因的 5′端启动子的结合,(如上文所讨论的,这个过程由于 σ 因子而更易进行)。开始,酶与双链 DNA **松散和可逆的结合**(loose binding),因为它还在寻找启动子序列——这是闭合的启动子复合物。当找到正确的启动子序列,启动子处的 DNA 就开始解链。酶和 DNA 之间的相互作用变得不可逆,发生**紧密结合**(tight binding),成为开放的启动子复合物。转录经常以 GTP 或 ATP 从头开始合成,与接下来的核苷酸不同,起始位置的 GTP 或 ATP 保留三磷酸基团。进行中的转录常常在一短寡核苷酸合成之后停止,聚合酶回到起始位置,称为流产起始(abortive initiation)。成功的起始经常在模板上形成 10 个或更多的核苷酸残基后开始。

这时聚合酶从启动子上游被清除。然后 σ 因子从核心酶中解离。

结构性调控——启动子结构　细菌启动子相对简单。大肠杆菌启动子有两个保守性基序直接与 σ 因子相互作用：-10 序列(Pribnow 盒)和 -35 序列(图 29.2)。每个基序的序列和它们相对位置对转录起始至关重要。-35 序列的突变影响起始结合的效率，而 -10 序列突变影响开放复合物形成的速度。这表明 -35 位置是 σ 因子的识别位点，σ 因子结合导致 -10 序列的解链，由于 -10 位置多个较弱的 A:T 配对解链变得容易进行(参见**热熔解**)。这两个位点之间的距离是双螺旋的一圈，使两个位点能与 σ 因子同时作用。两个位点之间距离改变的突变通过影响 σ 因子与两个结构域作用能力而改变起始效率 DNA 的拓扑学或构象的改变有同样的效果。转录起始点本身序列也会影响起始效率，这包括了三碱基结构 CAT 中间的嘌呤与转录本 +1 位是对应的。另外，启动子清理的效率受前五十个碱基或转录区内的一些碱基调节。

"缺省"的大肠杆菌转录起始效率有三个级别的变化，反映了启动子序列的变化。最弱的启动子完全没有 -35 序列，表达速度近乎零，需要另外的激活蛋白帮助 RNA 聚合物的结合(参见**代谢抑制**)。不同的大肠杆菌 σ 因子识别不同的序列，尽管两分结构看来是广泛保守的。在很多其他细菌中也有类似的启动子，但保守结构的相对位置有所不同。

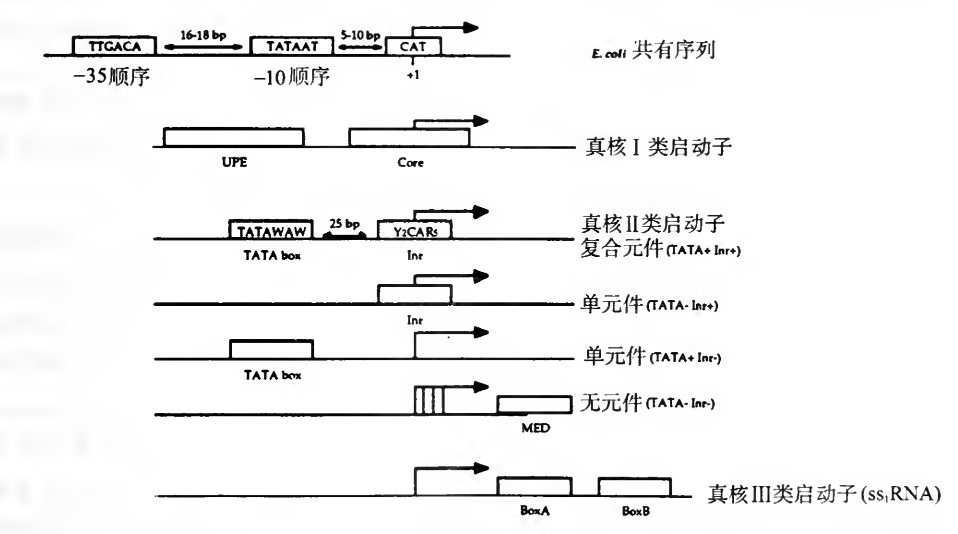


图 29.2　基本启动子结构。大肠杆菌共有启动子结构和真核生物Ⅰ型，Ⅱ型和Ⅲ型启动子。Ⅱ启动子根据有无 TATA 盒和起始区序列分为四类。Ⅲ型启动子是 5S rRNA 基因。其他Ⅲ型启动子类似与Ⅱ型启动子，可以有 TATA 盒和起始区元件。

29.3　真核生物转录起始——基本的结构成分

RNA 聚合酶　真核生物有三种依赖 DNA 的 RNA 聚合酶，每种负责一类基因的转录；它们的性质总结在表 29.2 中。它们根据对 α-鹅膏蕈碱(α-amanitin)敏感程度而区分，α-鹅膏毒环肽是一种霉菌毒素。真核生物 RNA 聚合酶的亚基组成还不十分清楚，它们都有 10 个或更多的亚基，其中一些为所有三种酶共有。每个聚合酶最大的亚基彼此同源，且与大肠杆菌的 α、β 和 β′亚基同源。RNA 聚合酶Ⅱ的类 β′亚基有一个柔性的 C 端结构

域,在转录起始和延伸中起重要作用(参见下文),并和剪接和多聚腺苷化因子结合(参

表 29.2　真核生物 RNA 聚合酶的一般特性

酶	功能	敏感性
RNA 聚合酶 I	转录 45S rRNA 前体,45S rRNA 前体是包含 5.8S,18S 和 28S rRNA 基因的多顺反子单位(Ⅰ型基因)	对α-鹅膏蕈碱不敏感,对放线菌素 D 敏感
RNA 聚合酶 Ⅱ	转录所有编码基因和大多数核内小分子 RNAs(Ⅱ型基因)	为α-鹅膏蕈碱抑制
RNA 聚合酶 Ⅲ	转录 tRNA 基因,5S rRNA 基因和编码 U6 snRNA 和不同的 scRNAs 基因(Ⅲ型基因)	根据种属对α-鹅膏蕈碱中等敏感

见第 27 章 RNA 加工)。真核生物中没有细菌 σ 因子的对应物,所以没有其他蛋白的帮助,就不能识别或结合启动子。

基本转录起始概述　　真核生物 RNA 聚合酶需要辅助蛋白因子,**转录起始因子**(transcription initiation factors, TIFs)或**一般转录因子**(general transcription factors, GTFs)帮助其识别启动子。GTFs 必须先与 DNA 结合形成复合物,帮助 RNA 聚合酶定位到 DNA 上的转录起始位点。聚合酶和进一步基本因子被征集到 DNA 上形成了**前起始复合物**。由于一个或多个 GTFs 的活性将 DNA 构象从封闭式转换成开放形式。GTFs 与 RNA 聚合酶本身组成了转录起始的**基本装置**(basal apparatus)。

真核生物三种 RNA 聚合酶相关的 GTFs 不同,但也含有相同的成分。其中最明显的就是 **TATA-结合蛋白(TBP)**,它对特定类型启动子有序列专一性作用(参见第 17 章核酸结合蛋白)。

组成型调控概述　　真核生物 RNA 聚合酶形成的起始复合物只有很弱的转录活性:在没有其他作用蛋白的情况下,它们组装缓慢且不稳定。起始效率通过结合在启动子附近顺式作用位点的蛋白因子而提高,对Ⅱ型基因来说结合在远处的增强子也可以增加转录起始效率。这些**转录因子**(transcription factor)提供形成起始复合物合适的环境,或与其中成分作用促使它们的结合或激活,这些因子称为**转录激活子**(transcriptional activators)。一些处在活性状态的转录因子的存在是受到调节的,这些因子参与了其靶基因对外源刺激反应的调节。另一些转录因子在所有细胞中有活力,因此对基因的组成型转录调控有帮助。细菌和真核生物转录起始显著的不同是细菌 RNA 聚合酶全酶不需要其他成分就能有效地进行转录起始,而真核生物的 RNA 聚合酶需要 GTFs 识别启动子和组建起始复合物,即使对强启动子还需要结构性调节因子来进行有效而足够稳定的转录。这反应了真核生物基因组普遍的负调控策略——大多数基因在大部分时间是以非活性形式存在,基因表达经常受选择性激活的调节,尽管还存在激活抑制的例子(参见沉默子,抑制结构域)。

RNA 聚合酶Ⅰ的转录起始　　真核生物基因组中只有一类Ⅰ型转录单位,编码核糖体 RNA 的多顺反子转录本。RNA 聚合酶的调控是很简单的。脊椎动物 RNA 聚合酶Ⅰ的启动子由两部分组成,包括转录起点附近的**核心启动子**(core promoter)和在起点 5′上游 100bp 左右的**上游控制元件**(upstream control element, UCE)。核心启动子本身对基本转录已经足够了,而 UCE 的存在大大加强了转录活性。一个称为 UBF1 的蛋白与两个元件的 GC 丰富区域结合,作为四聚体复合物 SL1 的识别位点,SL1 包含了 TATA 盒结合蛋

白 TBP 和其他 RNA 聚合酶定位于 DNA 所需的成分。UBF1 的活性受到与成视网膜细胞母瘤蛋白(retinoblastoma protein)相互作用的影响,使 rRNA 合成与细胞周期相协调。在一些低等真核生物中,SL1 的同源蛋白本身直接与 DNA 结合,而没有 UBF1。

RNA 聚合酶Ⅱ的转录起始　　RNA 聚合酶Ⅱ启动子变化最大,反映了控制很多编码蛋白基因的复杂表达图谱的调控元件的多样性。一般,这样的启动子可以分为位于转录起点的**基本启动子**(basal promoter)和包含结构性和可调节的元件的**上游启动子元件**(upstream promoter elements)。基本启动子变化很少,可以分成四类。而不同基因的上游启动子元件有不同的结构性和调控元件的构型,允许不同的转录调节(见下文)。基本启动子由位于转录起点的保守序列 YYCARR 的**起始子**(initiator, Inr)序列和叫做 TATA 盒的基序所组成[**TATA box**,也称为 **Goldberg-Hogness(GH)盒**,它位于 -25,保守序列为 TATAWAW]。这些基序分别类似与细菌起始位点(CAT)和 -10 序列(TATAAT),只是位置有所不同。起始区和 TATA 盒作为一般转录因子的识别位点,其功能是帮助起始复合物准确定位于转录起点处。一些没有 TATA 盒和起始区的Ⅱ型基因可能依靠下游的元件结合。这些启动子经常有**多个转录起始位点**(multiple start sites),反映了缺乏聚合酶的精确定位的功能。四类Ⅱ型基本启动子在图 29.2 中表示。目前没有证据表明任何特定的启动子类型一定比其他启动子更有效,保守元件的突变对不同基因的作用也是不同的。这说明启动子上下游的结构和它相关邻近的元件对转录起始的效率有很大影响。

转录起始的第一步是 TFⅡD 因子对 TATA 盒的识别,TFⅡD 包括了 TATA 结合蛋白 TBP 和一些 **TBP 结合因子**(TBP-associated factor, **TAFs**)。TBP 通过 DNA 的小沟与 TATA 盒专一性作用(见核酸结合蛋白)。TAFs 重要性有两个原因:一是它们可以识别核心启动子的非 TATA 元件,因而可能在只有起始子的启动子和裸露启动子的识别起重要作用(图 29.2);二是它们允许 TFⅡD 对转录激活子和沉默子作出反应。TAFs 在基本启动子的特殊组装决定了起始复合物与上游激活子如何作用,使基本启动子本身具备某些细胞类型专一性。例如淋巴专一性表达的末端脱氧核酸转移酶基因和骨髓专一性表达的 γ 干扰素,如果它们基本启动子结构改变就会失去调控。类似的果蝇乙醇脱氢酶基因在胚胎期和成体期之间启动子的改变也依赖于起始区元件的结构。

接下来更多的因子按顺序与启动子结合(图 29.3)。TFⅡA 第一个结合,并阻断如 DR1 等抑制 TFⅡD 的活性的因子的结合。然后 TFⅡB 结合并作为 TFⅡF 的桥梁分子。TFⅡF 携带 RNA 聚合酶Ⅱ进入复合物。TFⅡB 和 TFⅡF 也可能共同促进 RNA 聚合酶Ⅱ与起点间的相互作用。聚合酶帮助 TFⅡE 进入,TFⅡE 接着吸纳 TFⅡH(并刺激其活性)和 TFⅡJ。TFⅡH 对起始前复合物很重要:它有控制启动子解链的螺旋酶的活性和磷酸化 RNA 聚合酶Ⅱ C 端结构域(C-terminal domian, CTD)的激酶活性。磷酸化使聚合酶从起始复合物中解离,有利于启动子的清除,在启动子上留下一个残留的起始复合物。TFⅡH 还能磷酸化其他基本成分和涉及转录偶联修复(transcription-coupled repair,参阅)不同形式的蛋白。像细菌 RNA 聚合酶一样,RNA 聚合酶Ⅱ在启动子清除前可能会放弃起始。第一个插入的核苷酸经常是嘌呤,它立即被 mRNA 鸟嘌呤转移酶修饰产生一个帽结构(参阅)。RNA 聚合酶Ⅱ的转录起始位点有时也称为**帽位点**(cap site)。

RNA 聚合酶Ⅲ的转录起始　　RNA 聚合酶Ⅲ启动子分为三类,是 5S rRNA 和 tRNA 的

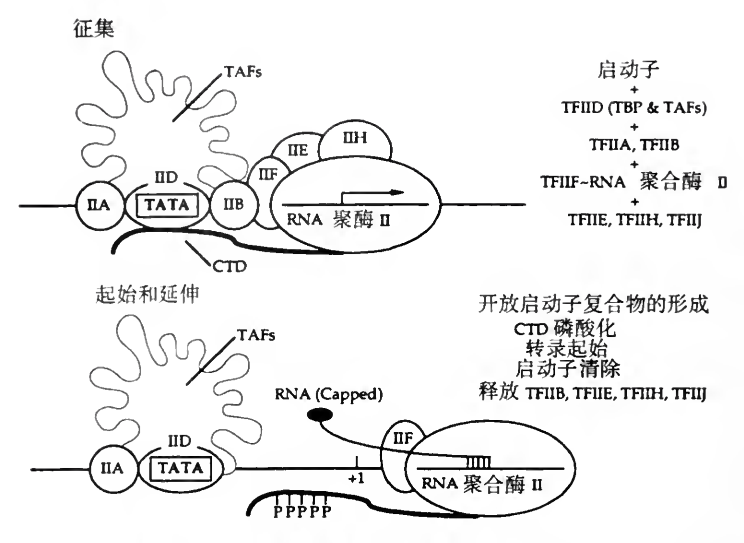


图 29.3 RNA 聚合酶Ⅱ启动子转录激活的阶段。

基因，和编码几个拥有**内部控制位点**(internal control sites，**ICS**)的 scRNAs 的基因，启动子位于转录起始的下游，并转录成为 rRNA(图 29.1)。5S rRNA 基因有Ⅰ**型启动子**，需要 GTFs(TFⅢA，TFⅢB 和 TFⅢC)的结合来定位 RNA 聚合酶。tRNA 基因有Ⅱ**型启动子**，只需要 TFⅢB 和 TFⅢC 的结合。Ⅲ**型启动子**控制 snRNA 基因的表达，组成类似经典的 RNA 聚合酶Ⅱ启动子，拥有 TATA 盒和类 RNA 聚合酶Ⅱ的调控基序；这些启动子比Ⅰ型和Ⅱ型启动子更复杂，可以引发细胞类型专一性的基因表达。RNA 聚合酶Ⅲ在Ⅲ型启动子的定位需要 TFⅢB 和其他辅助因子。TFⅢB 是各类 RNA 聚合酶Ⅲ启动子都需要的一个成分，它含有 TATA 结合蛋白并帮助 RNA 聚合酶进入起始复合物。在Ⅲ型启动子中，TFⅢB 直接与 DNA 作用，而在Ⅰ型和Ⅱ型中是通过蛋白-蛋白相互作用与其他 GTFs 作用，结合在前起始复合物中。这类似于 RNA 聚合酶Ⅱ启动子的含 TATA 盒与不含 TATA 盒的不同组建方式。

29.4 转录起始——调节成分

转录因子激活的基本原理 与基本启动子附近或远些的位点结合，控制起始复合物的组装、稳定性或活性的转录因子的概念，在上文真核生物组成型转录激活子中已经介绍。在原核生物和真核生物中，基本启动子周围都有短的序列区域，可以作为转录调节因子的结合位点。这些因子使用组成型转录因子相同的作用机制发挥作用，但可以是增加转录起始的正调节(**转录激活物**)或降低的负调节(**转录阻遏物**)。与组成型转录因子不同，它们的存在和活性受到如外在信号等因素的限制，使它们所调节的基因也受到类似的调节。

原则上，转录因子以四种方式发挥作用：

(1)通过与 DNA 结合和与基本起始装置的相互作用；

(2)通过与 DNA 结合并改变其结构和构象；

(3)间接的通过 DNA 结合并与其他调节因子作用影响其活性；

(4)通过蛋白-蛋白之间的相互作用(不与 DNA 结合),或直接作用基本起始物中的成分或间接与其他调节因子作用。

细菌和真核生物的转录因子根据它们结构域的特点分成一些家族。它们可以拥有几个结构域,主要是 DNA 结合结构域和激活(或抑制)结构域,经常也有二聚体结构域和与效应分子或信号传导成分作用的结构域。DNA 结合结构域不在这章中讨论(参见第 17 章核酸结合蛋白)。

一个特定基因的转录活性依赖于它本身不变的决定正常转录速度的内在的性质(调节元件的结构和真核生物中是否存在组成型激活物)和可变的外在因子(如是否存在转录调节激活物或阻遏物)。管家基因主要受启动子结构和组成型转录因子的控制,而可诱导或抑制的基因,以细胞类型的专一性或发育限制性方式表达的基因,也受转录调节因子的调控。

调节元件——启动子和增强子　顺式作用的调节元件经常在基本启动子附近发现,以有利于与 RNA 聚合酶(细菌中)或起始前复合物(真核生物)的局部相互作用。大多数细菌调节元件在启动子侧翼,而在真核生物中,组成型和转录调节因子经常结合在基本启动子的 5′端的上游区域。除了这些邻近位点,附加的顺式作用元件可以在距它们所调控基因上百甚至上千碱基距离处。这样远距离的元件如果是正调控则称为**增强子**(enhancer),负调控为**沉默子**(silencers),它们在真核生物中普遍存在,但在细菌中很少——如在大肠杆菌中增强子只与受 σ^{54}-因子调控的基因有关。增强子类似于这样的启动子——它们有结合调节和结构性转录因子的顺式作用元件,并与基本起始复合物作用调节转录起始。由于它们所处的位置,这样的作用通过间插的 DNA 环出介导,使增强子以不依赖位置和方向的方式发挥作用——因此它们可以位于上游、下游甚至所调节基因的内部。启动子调节元件在这方面受到更多的限制,启动子只能在一个方向起作用。一些作用只能发生在邻近的位置,所以启动子调节元件的位置对其功能很重要——大多数细菌操纵子和激活物/起始子属于这一类。其他相互作用需要起始复合物周围 DNA 弯曲和折叠的作用,这样的调节元件的位置的可变性较大——很多结合在启动子上游区域真核生物转录因子属于这一类。表 29.3 总结了转录中的顺式作用位点和反式作用因子。

表 29.3　调节细菌和真核生物转录起始的顺式作用元件和反式作用因子的总结。(+):正作用;(-):负作用

元件	定义和功能
细菌	
启动子(+)	在细菌中 RNA 聚合酶结合的位点。在大肠杆菌和很多其他细菌中,启动子由位于 -35 和 -10 的两个基序组成,被 RNA 聚合酶的 σ 因子识别。等同于真核生物的基本启动子。
起始子/激活子(+);操纵子(-)	细菌中的正/负作用调节元件,经常与启动子重叠,并影响 RNA 聚合酶的结合或活性。
增强子(+)	是远距离正作用的调节元件。结合的因子通过内部 DNA 的成环突出与启动子相作用。
真核生物	
启动子(+)	在真核生物中,基因邻近位置的控制位点包含两个成分:RNA 聚合酶结合的**基本启动子**(对等于细菌启动子)和**上游启动子**。后者包含**组成型元件和调节性元件**,一起控制转录起始。启动子可延伸到基因较远距离,但以方向依赖的方式作用。大部分启动子在基因紧接着的 5′端,但 RNA 聚合酶Ⅲ启动子可以位于转录区域内(内部控制位点)。

续表 29.3

元件	定义和功能
增强子(+);沉默子(-)	增强子是真核生物中的正作用调控位点,经常包含组成型和可调节的元件,从非常远的距离(<50kb)作用,可以使它的靶基因的转录活性增加达 1000 倍。增强子不依赖方向和位置,因为它通过使内部 DNA 成环突出而与基本转录装置作用,但没有启动子活性,这是因为缺乏基本的启动子成分。增强子可以包含有功能的成簇的结合位点群,称为**增强子单元**(enhancons)。它们可以调节多个启动子,在有些系统中引起增强子的竞争。沉默子是与增强子作用特点相同的负调控作用位点。
上游激活位点(+);上游抑制位点(-)(UAS,URS)	酵母中类似于增强子或沉默子的位点,以类似的方式发挥作用,但不能调节所控制基因的下游区域。
基因座控制区域	通过建立一个开放的染色质结构域控制基因表达的远距离调节元件。减弱染色质沉默引起的顺式抑制,允许位置非依赖性,拷贝数依赖性的转基因表达,但不以经典增强子的方式增强瞬时转染测试中的转录活性。
转录因子(TFs)	**定义和功能**
细菌	
σ因子	识别启动子的 RNA 聚合酶亚基。
转录激活物/阻遏物	结合于起始子/激活物和操纵子位点,影响 RNA 聚合酶结合/活性的蛋白。
真核生物	
基本 TF/一般 TF,转录起始因子(TIF)	例如转录因子 TFⅡD,是基本启动子处转录前起始复合物的一部分,为 RNA 聚合酶结合或活性所需。根据 TFab 系统命名,a 是同性质 RNA 聚合酶的序号,b 按照被发现的顺序。因此 TFⅡD 是 RNA 聚合酶Ⅱ起始复合物中第四个被发现的转录因子。
组成型 TF	例如转录因子 Sp1,在所有细胞中存在,对基因表达有普遍的正调控作用,与上游的启动子和增强子中的结构性元件结合。
转录激活物/阻遏物	例如转录因子 MyoD1,激活形式的存在是自身调节的。与上游启动子和增强子的调节元件结合,帮助调节基因的表达。

基因座控制区　　基因座控制区(locus control region, **LCR**)是真核生物的顺式作用元件,经常位于所调控基因的较远距离,这对转录活性有重要作用,因为它建立了独立的染色质结构域。LCR 的重要性在分析珠蛋白在转基因小鼠中的表达实验中得到证实,通过这类实验,产生了定义 LCR 的方法。β-珠蛋白基因可以通过转染导入培养的红细胞中,并以附加体的形式维持短时间。如果结构中包含β-珠蛋白的启动子和增强子,β-珠蛋白的表达量就很高。相反,同样的结构导入小鼠的基因组中,证明即使在红细胞中也只有很少的活性,这是因为它被染色体附近区域所抑制。如果此结构和基因座控制区域连接,顺式抑制作用就会减弱,使β-珠蛋白的基因能够在转基因小鼠的红细胞系中表达。LCR 不仅使整合基因可进行位置非依赖的表达,而且转基因表达水平也直接与拷贝数成正比。LCR 对转染细胞的转录没有影响,它并不仅是一个增强子。

像启动子和增强子那样,珠蛋白的 LCR 包含了很多转录因子的结合位点,一些是不变的,一些是红细胞专一性的。因为它不影响培养细胞的转录速度,但在体内是有影响的,所以 LCR 被认为是通过改变染色质结构起作用(附加体维持 DNA 不被包装成染色质,没有顺式抑制的压力)。LCR 如何调节它的作用还不得而知,但在某些方面是和与β-珠蛋白基因族基因暂时调节的途径相关(参见文框 29.1 的进一步讨论)。

细菌中与 RNA 聚合酶作用的转录因子　　很多转录因子通过与起始复合物中的一个成分作用,直接影响起始而发挥功能。在细菌中,这样的因子与 RNA 聚合酶本身作用。在大肠杆菌中,一个简单的转录抑制的模型是 *lac* 操纵子,在这个模型中,Lac 阻遏物结合在与启动子重叠的主要操纵基因位点,直接阻断了 RNA 聚合酶进入启动子。另外,Lac 阻遏物和 RNA 聚合酶都与 DNA 结合,但阻遏物和聚合酶的相互作用阻止了开放起始复合物的形成。这是其他操纵子共同使用的策略。在 *aroH* 和 *trp* 操纵子中启动子和操纵基因有明显的重叠,说明阻断/置换是抑制的可能机制,在代谢抑制的条件下 CAP 和 RNA 聚合酶结合在附近位置负责 *lac* 和 *gal* 操纵子的转录激活(见文框 29.2)。启动子处的 RNA 聚合酶和结合在增强子的转录因子通过内部 DNA 的成环突出可能发生正向的相互作用。结合在 *glnA* 基因启动子的 RNA 聚合酶形成一闭合的启动子复合物,但开放的启动子复合物形成需要结合在 150bp 远处增强子的 NtrC 转录激活子。激活物和 RNA 聚合酶的直接物理作用为 DNA 解链所需。在特定的情况下,转录因子的活性还可以影响 RNA 聚合酶对启动子的专一性。枯草杆菌孢子形成的过程受几个 σ 因子和控制启动子专一性的转录因子合成的调节(参见文框 6.2)。

调节 DNA 结构的细菌转录因子　　在原核生物中,DNA 的拓扑学特点会影响转录。编码 DNA *回旋酶*(参阅)的突变或转座子的插入会使隐藏操纵子去抑制,被认为是改变了 DNA 的拓扑结构,使其能被 RNA 聚合酶所识别。更有一些转录阻遏物通过分离启动子为拓扑结构上紧密的 DNA 环,抑制开放起始复合物形成起作用。大肠杆菌 *gal* 操纵子的双启动子为操纵基因从两侧包围,它们协同结合 Gal 的阻遏物。每个位点上结合的阻遏物单体结合产生二聚体,分离含启动子的 114bp 的 DNA 环,使其双链不能解链。由于结合在主要操纵基因和两个次要操纵基因中的一个上的 Lac 阻遏物的相互作用,使 lac 启动子也形成类似的环。

真核生物中与前起始复合物相互作用的转录因子　　在真核生物中,大部分直接与前起始复合物作用的转录因子是转录激活物。转录因子结构域分析显示**激活结构域**(与转录基本装置作用,帮助转录激活)可以分为为数不多的家族(见文框 29.3,参见*结构域交换*),如 DNA 结合结构域(参见第 17 章核酸结合蛋白)。酸性激活结构域在不同物种中是高度可互换的,说明特定激活家族的所有转录因子都与同样的、高度保守的基本转录装置成分作用。**压制**(squelching)的现象支持了上述结论——当转录因子过度表达,不仅激活它所有的靶基因,而且通过蛋白-蛋白的相互作用,抑制了非正常调节的基因的基本转录装置,引起全局性的下降调节。所有三种类型的激活结构域都刺激 TFIIB 进入起始复合物,酸性激活结构域诱导 TFIIB 的构象改变,增加其对 RNA 聚合酶的亲和力。激活结构域也影响了基本转录装置中的一些成分的活性,包括 RNA 聚合酶本身。在酵母和人中发现了多组份的**调节介质**(mediator),它装配在 RNA 聚合酶 Ⅱ 的 C 端结构域,传导来自转录激活子的信号。TATA 结合蛋白也是转录激活子的靶蛋白,转录激活通过与 TAFs 的作用而调节。不同的转录因子与不同的 TAFs 作用,不同的 TAFs 以不同的构型在启动子上排列。这使基本启动子通过指示提供的 TAFs 与特殊的上游转录激活子作用,能够控制细胞-类型-专一性的基因表达(文框 29.3)。已知一些转录因子是起多信号的综合分子的作用:如 CBP 是通过 DNA 成环突出作用的,它是具有 10 个不同转录因子

结合位点的巨大蛋白。转录激活子也可以与几个其他 GTFs 作用,包括 TFIIH。这样多个因子作用于前起始复合物的多个靶成分而调节转录的组装速度、稳定性和活性。

与大多数细菌阻遏物直接作用于 DNA 聚合酶不同,大多数真核生物转录阻遏物通过抑制激活物的功能起作用(见下文)。一些直接作用的阻遏物已经分离得到,它们调节功能的方式与直接与起始复合物作用的激活物不同。这些因子通过蛋白-蛋白相互结合作用,如 DR1 抑制前起始复合物的组建,而其他在启动子或远距离的沉默子元件上有隐藏的顺式作用位点,抑制前起始复合物的组装或激活(如果蝇配对-原则蛋白 Even-skiped 和哺乳动物的甲状腺激素受体在没有其相应配基的情况时)。

改变 DNA 结构的真核生物转录因子 在真核生物启动子中,一些转录因子的功能是置换核小体,产生 DNase I 高度敏感位点(参阅)。这是一种转录激活的直接途径,通过允许其他转录因子与附近原先为核小体占据的位点结合,或是有利于启动子处染色质重排,减弱由染色质高级结构引起的转录抑制。这种因子的例子包括果蝇中的普遍存在的 GAGA 因子及酵母和脊椎动物中与 RNA 聚合酶 II C 端结构域相联系的 SWI/SNF 复合物(参见组蛋白乙酰化)。一些结合在增强子和沉默子的转录因子也有类似作用机制,通过大范围连续的染色质结构的改变起作用。例如酵母交配型盒 HMLα 和 HMRα(参见交配-类型转换)具有它们自身的启动子,但通过位于几千碱基远处的沉默子的异染色质化保持一种非活性状态(受保护免受内切酶作用)。称为 RAP1 的蛋白识别这些沉默子(和端粒沉默子位点),并征集四个 SIR 蛋白(沉默信号调节子)中的 SIR3 和 SIR4,它们与组蛋白 H3 和 H4 作用,并可能多聚化成为异染色质过程的一部分;它们也可能将沉默区域与核基质联系。

一些真核生物的转录因子通过使 DNA 弯曲发挥功能,这样有助于其他成分的作用。很多是 HMG 蛋白,普遍存在的 HMGI/Y 和 HMG14/17 在染色质结构中起重要作用(见第 3 章染色质)。像 Sry(控制哺乳动物性别决定)、Sox-2、Sox-3 和 Sox-11(与神经分化有关),以及 Lef-1(控制淋巴细胞专一性基因表达)的 HMG -盒转录因子能够将 DNA 组织成特殊的三维构象,在这样的结构中不同的转录激活物有效地相互作用[这样的结构称为**增强体**(enhanceosome)]。这提供了另一个启动子中顺式作用位点对它的活性是如何必需的例子——重新使 DNA 弯曲复位可以破坏它们包装结合蛋白的次序,导致无效的相互作用并失去转录活性[参见 V(D)J 重组]。

转录因子相互作用调节转录活性 大多数真核生物转录阻遏物是通过妨碍激活子的功能被动地起作用,而不是直接与转录前起始复合物作用。阻遏物一种普遍性作用机制是占据激活物的结合位点或与其重叠在附近的位置,这样从空间上阻止激活物的结合。在这种情况下,所调节基因的命运取决于激活物和阻遏物结合的竞争和每个转录因子对相应位点的相对亲和力。一个弱结合的激活物可以被高亲和力的阻遏物所取代。果蝇 *even skipped* 基因提供了一个通过竞争结合位点调节转录的例子,它是通过不同 gap 蛋白之间竞争上游结合位点调节转录。在其他情况下,正负作用蛋白竞争与转录激活子的作用,决定激活子是否能与 DNA 结合及结合后是否能激活转录。这种情况最简单的例子是螺旋-环-螺旋和亮氨酸拉链家族的转录因子二聚体作用。二聚体可以包含两个活性状态单体(DNA 结合和转录激活)或者是一个活性单体和一个非活性单体而产生不能与

DNA 结合的非活性异源二聚体。这方面的例子包括缺乏 DNA 结合结构域的 Id 对肌源性 bHLH 蛋白的抑制,和活性 bHLH 蛋白(如 sisterless-a)与抑制 HLH 蛋白(如 Deadpan)之间的平衡对果蝇性别决定(参阅)的控制。

非直接抑制的其他机制不以竞争为基础,增加激活物水平没有效果。这在**淬灭**(quenching)的现象中可以观察到,阻遏物结合于 DNA 一定位点上阻碍一个已经结合激活子的活性,经常是阻断其进入基本转录装置或屏蔽它的激活结构域。从此延伸,激活子也可以偶尔以这种方式作用,通过阻断阻遏物的抑制结构域而"去淬灭"。

转录因子的调节 只有在非常特殊的环境下,基因周围的顺式作用元件才能被调节(参见周期转换、交配型转换)。转录调控经常是依赖于所提供的活性转录因子可调节性。原则上,转录因子可以两种方式被调节——通过控制它们的量和活性。在细菌和真核生物中两种方式都有,但在细菌中后者是主要的:大多数转录因子的合成是组成型的,在翻译后调节。

转录因子从新合成的调节方式是经济的(资源没有浪费在合成不需要的蛋白上),但在细胞质中决定合成因子和因子出现正确的水平之间存在着延迟。所以在转录活性需要长期改变时,转录因子的重新合成机制是有用的,但在细胞需要对外刺激作出快速和瞬时反应时则不然。因此在多细胞器官发育过程和维持已分化的状态时是主要的。在细菌中,重新合成用来调节长期对环境的基本反应,如枯草杆菌的出芽。

转录因子活性的调节解决了调节链中的问题 - 如果一个转录因子本身的转录水平上受到调节,那么一定存在另一个转录因子控制这种调节,调节的问题被沿着链简单地转移到下一个环节。活性的调节也有利于对外在刺激正确反应,特别是当转录因子的活性直接与刺激相关。在最极端的例子中,刺激本身就是转录因子:嗜中性细胞对细菌感染反应,释放乳铁蛋白(lactoferrin),表达量增加并与 DNA 结合,像转录激活子一样,诱导很多控制反抗细菌感染基因的转录。与之近似是刺激变成为转录因子的**变构制剂**(allosteric modulator)。这在细菌中相当普遍,细菌在环境中的代谢控制着转录因子,转录因子调节控制它们合成或降解的基因。在 *lac* 操纵子中,乳糖的代谢产物别乳糖与 Lac 阻遏物结合并将其灭活,使编码利用乳糖的酶的基因得以转录(文框 29.2)。在哺乳动物中与其类似,类固醇激素可以作为它们受体的调节因子,它们的受体也是转录因子。再进一步就产生了信号传导途径——刺激作为信号,与受体结合,直接或间接激活转录因子。JAK-STAT 的信号传导途径是直接的 - 细胞因子的信号与细胞表面的受体结合,并激活一个细胞内 JAK 激酶,JAK 激酶磷酸化并激活 STAT 家族的转录因子。其他信号在到达它们的转录因子前要经过更长的途径,涉及很多成分(见第 28 章信号传导)。

信号传导途径提供了另一种控制转录因子活性的范例:共价修饰。真核生物的很多转录因子是通过磷酸化激活(或灭活),这样可以影响它们与 DNA 结合的能力(如 c-Jun)和激活转录的能力(如 c-Fos, CREM)和形成二聚体的能力(如 STATs)以及与它们的调节蛋白作用的能力(如 CREB 当被蛋白激酶 A 磷酸化后与 CREB 结合蛋白 CBP 作用)。在细菌信号传导途径中可以看到同样的机制,如与 *glnA* 基因上游增强子结合的 NtrC 必须被磷酸化才能结合 ;它被对环境氮源作出反应的 Ntr A 激酶所磷酸化。或者,信号途径的靶蛋白也可以是转录因子活性的一种抑制蛋白。一个明显的例子是 NF-κB 被磷酸化激活,对靶蛋白(抑制蛋白 I-κB)进行降解。在这种情况下,此因子同时受到激活和剂量

二种方式的调节。I-κB 屏蔽 NF-κB 的核定位信号，通过将其分散到细胞的其他空间阻止与 DNA 的结合。

组合和前后关系依赖性的调节 即使知道单个转录因子作用的效果，但两个不同的转录因子同时作用一个基因的效果是不可预计的。这种情况的一个例子是**协同作用**(synergism)，由两个转录因子共同驱动的基因转录活性要大于两个单独作用因子的效果的相加。转录因子 Pit－1 与雌激素受体协同作用促乳素基因，可能是通过增加基本转录装置不同成分的活性、募集稳定性起作用。在另一种情况，两个转录因子在独立作用时是激活基因，共同作用时却引起基因的沉默。这种关系依赖的调节可能根据转录因子的丰度或 DNA 上是否有特殊的调节位点。p53 因子提供了一个多水平关系依赖调节的例子：如果启动子提供识别位点时，p53 是基因表达的一个强激活物，但如果没有识别位点，p53 与 TFIID 直接结合抑制转录。除此以外，WT1 蛋白单独作用是转录激活物，当与 p53 结合时是转录阻遏物。结合的特异性也可因为不同的二聚化而改变(这发生在果蝇的同源域转录因子与辅助因子作用时与不同的位点结合)。

29.5 细菌和真核生物转录调节的策略

细菌转录调节策略的特点 细菌生活在一个动态的环境中，生存需要对代谢物和其他分子水平的变化作出快速和正确的反应。为了将这些反应线性化，细菌基因组经常将功能相近的基因(那些编码共同代谢途径的酶)组织成**操纵子**，使其处于共同的顺式调节(参见基因结构和作图)。操纵子的基因转录成多顺反子，在转录水平作为一个单元激活或抑制。这依赖于细菌核糖体进入内部起始位点的能力(参见蛋白合成)。转录调节主要发生在两个阶段：调节转录起始的**启动子调节**(promoter control)和调节终止的**衰减调节**(attenuator control)(衰减调节在这章的后面讨论)。

转录的快速诱导和抑制由细菌中启动子调节的主要形式——**别构调节回路**(allosteric control circuits)所协调，转录因子的合成是不变的，但它们的活性受与环境中小的效应分子(effector molecules)的作用影响，效应分子的存在反映了对所调节的基因产物的需要。一个例子是 *lac* 操纵子(文框 29.2)。编码转录因子的基因可以受自身调节(如受它们自身产物调节)或不受调节(转录起始的速率依赖于启动子的结构)。有四种类型的别构回路，在图 29.4 中表示。

细菌的长期反应经常由合成新的识别特定启动子结构的 RNA 聚合酶成分调节。在大肠杆菌中，对热休克和氮源缺乏的反应是新 σ 因子的合成；在枯草杆菌中，孢子生成受 σ 因子级联调节。

因为细菌基因排列紧密，可以应用一种转录抑制形式，即一个启动子的转录可以阻断另一个启动子的转录。这称为**逆向转录**(countertranscription)，基因转录产生的反义 RNA 在另一个方向上称为**逆向转录 RNA**。逆向转录在 λ 噬菌体控制裂解和溶菌广泛应用；它可能通过阻止正常启动子的转录起始、延伸、形成三聚体结构或反义效应起作用。大肠杆菌基因组的例子是 CAP 蛋白，它通过激活转录单位中的逆向启动子的转录抑制自身基因的转录。逆向转录在真核生物中偶尔也有出现(参见亲本印迹)。

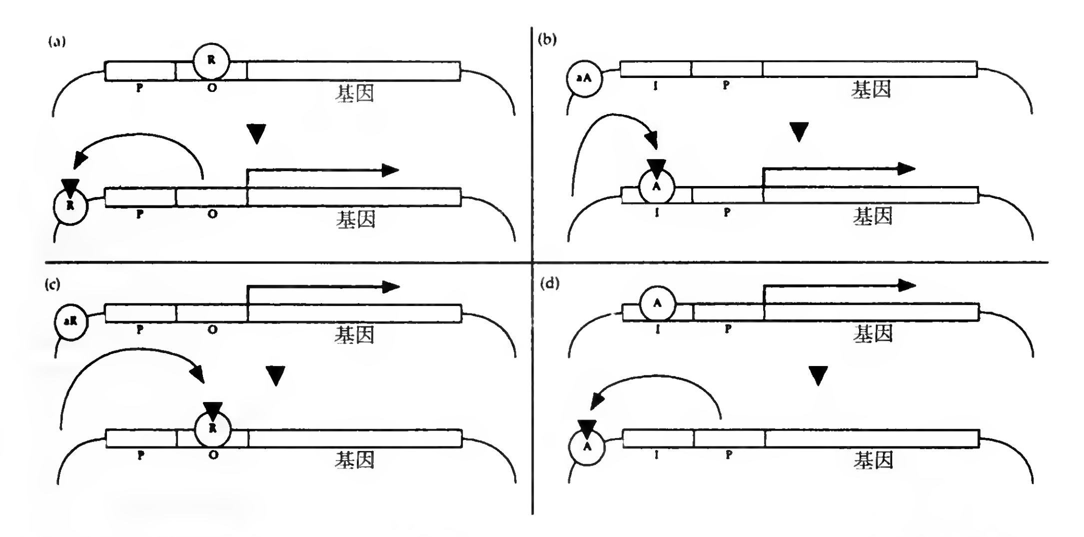


图 29.4 细菌中别构调节回路。(a)**负诱导调节**:基因在缺省状态下是关闭的,因为调节元件(**操纵基因**)被负作用因子(**抑制蛋白**)所结合。加入一个小效应分子(**诱导物**)引起阻遏物构象的改变,使阻遏物从操纵基因上释放出来,非特异地与 DNA 结合。基因变得去抑制和转录活性。例如:*lac* 操纵子,在乳糖缺省时被抑制,被乳糖的类似物别乳糖所诱导。(b)**正诱导调节**:基因在缺省状态是关闭的,因为正调节因子(**激活子蛋白**)不能与调节元件(起始子或激活子)结合。在这种状态下,激活子蛋白称为**非激活物**(apoactivator),与 DNA 的结合是非特异性的。加入一小效应分子(**辅活化物**)引起非激活物构象的变化,使其能特异性与起始元件结合并转录激活。例如:*lac* 操纵子的代谢抑制——CAP 蛋白是非激活子,只有在 cAMP 存在的情况下与 DNA 结合帮助起始。(c)**负抑制调节**:基因在缺省状态下是开放的,因为负调节因子(**抑制蛋白**)不能与操纵基因结合。在这种状态下它作为**非阻遏物**(aporepressor),而且非特异性地与 DNA 结合。加入小效应分子(**辅阻遏物**)引起非阻遏物的构象改变使它与操纵基因特异性结合抑制转录起始。例如:编码涉及色氨酸合成酶的 *trp* 操纵子。色氨酸本身是辅阻遏物,帮助关闭本身合成(负反馈)。Trp 阻遏物也抑制本身基因(**自身调节**)。(d)正抑制调节:基因在缺省状态下是开放的,因为因子(激活子蛋白)与调节元件(起始子)结合。加入小效应分子(**阻遏物**)引起激活物的构象改变,使它从起始子中释放,与 DNA 非特异性结合。基因就被阻遏,不进行转录。

真核生物调节策略的特点 在真核生物中,在缺省状态下所有基因都是非活性的,因为它们表达需要组成型转录激活物。真核生物中对外界刺激的快速反应主要受正可诱导调节介导,如通过潜在转录激活物的激活。细菌细胞中主要是变构调节,但在整合很多信号途径的真核生物中对转录因子本身或与之作用的蛋白的共价修饰是真核生物可诱导系统调节的基本机制。这样的转录因子与靶基因上所谓的**反应元件**(response elements)结合(表 29.4)。长期的过程,如多细胞器官的分化,主要由转录因子的重新合成调控。但细胞类型专一性的和发育相关的转录因子也通过与靶基因的启动子和增强子的反应元件结合发挥功能,这方面的例子在表 29.4 中列出。

细菌和真核生物调节策略的主要差别是真核生物缺乏操纵子。除了少数例外,所有的真核生物的转录本都是单顺反子,反映了真核生物核糖体不能从内部位点起始蛋白合成。一些真核生物基因成簇排列,处于共同的顺式调节(参见基因座控制区域、同源异形基因、亲本印迹)和远距离与多个启动子作用的增强子调节,使基因处于共调节(这在免疫球蛋白基因座有两个受 κ 增强子影响的启动子)。但在这些例子中,这些基因并不是一起转录的。因此真核生物需要一起调节的基因经常是分散的,但处于共同的反式调节。拥有共同的调节元件的基因可以被一起激活,但有重叠元件的基因可以被单独激活,反映了

表 29.4 一些真核生物基因上游涉及组成型、诱导、细胞或品系专一性和发育的调节基序以及相应的调节因子

共同序列	元件名称	转录因子	注解
组成型位点			**分类**
GCCAATCT	CAAT 盒	CTF/NF-1 CP 家族 C/EBP(bZIP)	普遍存在的 普遍存在的 普遍存在的，在肝中水平较高
GGG CGG	GC 盒	Sp1(zf)	普遍存在的
ATGCAAAT	八聚体(Octamer)	Oct1(h) Oct2(h)	普遍存在的 B-淋巴细胞
反应元件			**刺激**
CNNGAANNTCCNNG	HSE	HSF(bZIP)	热休克
CCATATTAGG	SRE	SRF	血清中的生长因子
TTNCNNNAA	IGRE	STAT1	干扰素-γ
细胞类型专一性元件			**细胞类型**
GATA		GATA-1(zf)	红细胞
ATATTCAT		Pit-1(POU)	垂体
CANNG	E-盒	MyoD1(bHLH)	成肌细胞
GGGACTTTCC	κB 位点	NF-κB	淋巴细胞
TTYAGNACCRCGGASAGNRCC	NRSE	NRSF/NEST(zf)	非神经细胞
发育调节位点			**发育体系**
TCCTAATCCC		Bicoid(h)	果蝇 AP 轴特异
GCGGGGGCC		Krox-20(zf)	脊椎动物后脑发育
TAATAATAATAATAA		Antennapedia(h)	果蝇同源异型基因
TCAATTAAATGA		Fushi tarazu(h)	果蝇配对法则基因

转录因子根据 DNA 结合/二聚化结构域分类(bHLH，碱性螺旋-环-螺旋；bZIP，碱性亮氨酸拉链；H，同源异形结构域；POU，POU 结构域；zf，锌指结构)。C/EBP(CAAT/增强子结合蛋白)与两个特定的位点结合，是 CAAT 盒和增强子核心序列 TGTGGWWWG。缩写：CTF/NF1，CAAT 转录因子/核因子 1；CP：CAAT-结合蛋白；HSE/F 热休克元件/因子；SRE/F：血清反应元件/因子；IGRE：干扰素-γ 反应元件；STAT：信号传导子和转录激活物；NRSE/F：神经局限的沉默元件/因子。

转录控制的弹性。在真核生物中处于共同反式调节的一组基因称为**基因群**(gene battery)[在细菌中的对应物称为调节子(regulon)]。一个基因群或调节子由受同一调节因子控制的分散基因和操纵子组成。有三种情况中真核生物也有多顺反子：①不翻译的核糖体 RNA 基因，这些基因不被翻译，因此也不使用核糖体；②线虫中双顺反子转录本通过选择性反式剪接产生含有上游加帽的开放阅读框架的成熟转录本；③多顺反子的细小核糖核酸病毒基因组。细小核糖核酸病毒有内部核糖体进入位点(internal ribosome entry sites，参阅)使核糖体能和每个开放阅读框架的上游结合(这样元件在基因扩增被广泛应用；参见基因敲除，报告基因)。

29.6 转录的延伸和终止

延伸和聚合酶的进行性　细菌和真核生物延伸的生化和动力学相似。延伸不是光滑的:RNA 聚合酶遇到障碍物(如二级结构)会停顿,这样会导致**受阻**(酶与转录本末端作用松散)或**终止**(酶与模板作用也松散)。在真核生物中,有一些在延伸过程与酶结合的**延伸因子**(elongation factors)抑制停顿[如延伸子(elongin)]和防止受阻(如 TEFb,TFIIS)。TFIIS(也称为 SII)可能是通过对新生转录本的剪切,在酶的活性位点处放置一个新的 3′端来阻止停顿(当酶停顿时新生链可以继续延伸,这样在活性位点外放置一个反应性 3′羟基)。尽管对大多数真核生物基因延伸是不变的,停顿和终止可以被调节,在骨髓细胞中 c-Myc 蛋白水平主要在转录延伸和终止的阶段被调节。在成熟的粒细胞中,全长 c-*myc* mRNA 被合成和翻译,而在未分化的细胞中,转录在第一个外显子被终止。停顿也调节了 HIV 基因组的转录,在感染早期产生少量,但大多数是被截短的转录本,而接着产生全长的转录本。在这个例子,调节是由病毒编码的与 RNA 上顺式元件 Tar 作用的反式作用因子 Tat 所介导。Tat 增加了转录起始的频率,并减少了聚合酶停顿。

细菌和真核生物基因大小的差异表明完成细菌中的转录较快而真核生物较大基因的转录需要几个小时——人类目前最大的基因,2.5Mbp 的肌营养不良蛋白(dystrophin)需要 16 个小时。最近有证据表明延伸的 RNA 聚合酶 II 在 DNA 上移动时,与多腺苷酸化装置和剪接体二者相结合,这两种复合物通过 C 末端结构域相互作用。这样肌营养不良蛋白和其他一些基因剪接和转录是同时进行的。

细菌中的转录终止　在细菌中,转录在分散的序列——**终止子**(terminators,**t**)终止,涉及的两种机制都是对转录本而不是基因本身的信号反应。

最常见的是**内部终止**(inteinsic termination)(**ρ-非依赖性的终止**,ρ-independent termination):转录本应用特殊的二级结构,引起 RNA 聚合酶停顿,DNA-RNA 核酸杂交链解离。一个普通的终止基序是后面 GC-丰富反向重复序列紧跟多聚 U 序列:GC 丰富的回文结构形成发夹,阻止延伸复合物,使 DNA 和 RNA 留在结合力弱的 A:U 碱基配对处,使 RNA 聚合酶从模板上解离。到目前还不清楚解离具体是如何发生的。它并不像最初设想的是由多 A:U 碱基配对弱本身引起的。另一种模型显示 RNA 聚合酶与其产物之间的相互作用起了决定作用:弱碱基配对可能有利于新生 RNA 末端从模板上被取代,并从酶的活性位点处松脱。这样就引起停顿,在这点上聚合酶所采用的构象允许发夹环进行取代。

另一种终止机制,**ρ-依赖性终止**(ρ-dependent termination)在细菌染色体中很少,而在噬菌体中广泛存在。它需要一个称为 ρ(**rho**)的蛋白与 RNA 聚合酶的直接作用,使 RNA 从 DNA 模板上脱离,可能是通过在酶的活性位点处挤入 DNA:RNA 的杂交链。ρ 与转录本的特殊位置结合,而不是与杂交链,特殊位置可能是大约 50～100 核苷酸长度的胞嘧啶富集而鸟嘌呤缺乏的区域。ρ 因为被认为是与延伸复合物一起沿着转录本移动,直到将转录本从 DNA 上释放出来,尽管目前没有直接的证据表明 ρ 蛋白沿着 RNA 移动,但有 ρ 因子维持在起始的结合位置处的证据。在**热追赶模型**(hot pursuit model)中,延伸着的转录本比 ρ 因子更快,ρ 不能紧跟转录泡,除非聚合酶发生停顿。目前还不知道是

如何导致在ρ依赖终止子停顿。在一些例子中发现了发夹结构(λ噬菌体的 TR_1 终止子和大肠杆菌的 his 操纵子),但大多数停顿现象在去除这些结构后仍存在。抗终止(antitermination,参阅)通常在ρ依赖的终止子处发生。

真核生物的转录终止 真核生物的转录终止研究得很少。RNA 聚合酶Ⅰ的终止发生在成熟 rRNA 3′端大约 1kb 左右的位置,与特殊的顺式作用元件识别有关。RNA 聚合酶Ⅲ的转录终止发生在类似细菌ρ-非依赖终止子的位置(紧跟多聚尿嘧啶的 GC 丰富区域),GC 序列看来并不形成二级结构,终止发生在第二个尿嘧啶核苷酸,反映了腺嘌呤:尿嘧啶的弱配对作用在解离中并不起关键作用。大多数 RNA 聚合酶Ⅱ的转录本通过 3′剪切和多聚腺苷化加工(参见第 27 章 RNA 加工),所以内部的终止反应并不清楚;转录可能比基因的 3′末端更远。然而最近有证据表明剪切和多聚腺苷化装置是与延伸中的 RNA 聚合酶相结合的,显示剪切和多聚腺苷化可能与停顿以及 RNA 聚合酶Ⅱ从模板上的解离是相关联的。

细菌中转录终止的调节 细菌能在终止水平上对转录进行正调节[通过刺激对终止位点的通读——称为**抗终止**(antitermination)的过程]或者负调节[通过诱导不成熟的终止——称为**衰减**(attenuation)的过程],这两种过程都被用来在操纵子结构中从多顺反子 mRNA 加入或减少阅读框架。

抗终止发生在λ噬菌体基因表达的短暂控制和大肠杆菌 rRNA 操纵子中。RNA 聚合酶在到达终止区前经过一个**抗终止位点**(antitermination)被修饰,抗终止位点结合有**抗终止蛋白**,如λ噬菌体 N 或 Q 蛋白(参见病毒和亚病毒)。对抗终止蛋白的作用似乎没有关键的位点:噬菌体的抗终止位点 *nutL*, *nutR* 和 *qut* 分别位于启动子、终止子和启动子上游。修饰的精确情况还不是很清楚。在 Q 终止体系中,RNA 聚合酶启动子处停顿时与 Q 结合,并携带着 Q 沿着模板移动。类似的 N 蛋白与 RNA 聚合酶和其他蛋白 NusA、NusG、NusB 和核糖体蛋白 S10(NusE)形成紧密的复合物。它们可能通过控制延伸的动力学性质,即作为停顿抑制子直接作用,也可能通过阻断停留或直接控制模板结构特异性地与终止装置作用。

衰减是细菌控制几个与氨基酸合成有关操纵子,以及在大肠杆菌中编码嘌呤合成的 *pyrBI* 操纵子的表达调节机制。在衰减控制中,转录终止通过蛋白合成效率(在氨基酸操纵子)或转录本身(在 *pyrBI* 操纵子)控制,使转录被这两个重要反应的底物有效调节。

对氨基酸操纵子而言,当缺乏特定的氨基酸使核糖体延迟,引起位于主要开放阅读框上游的编码**前导肽**(leader peptide)短开放阅读框架翻译减少。受延迟核糖体影响,新生 mRNA 采取有利于延伸的结构,促进生物合成酶的合成。相反,当特定氨基酸充足时,前导肽翻译增加,mRNA 就采取有内部终止子的结构,阻止不需要的酶的合成。前导肽的 ORF 的特征是含几个串联的氨基酸密码子,它们的合成受下游基因调控,例如在 *his* 操纵子前导肽有九个组氨酸密码子。在缺乏组氨基时,核糖体会在这个衰减子位点处延迟而转录延伸继续,但如果有足够的组氨酸,前导肽被翻译,转录被终止;组氨酸生物合成有关酶只有在细胞检测到组氨酸缺乏时才合成。

在 *pyrBI* 操纵子中,如果 UTP 有限的话,转录在含多个串联腺嘌呤的区域减慢,使核糖体抑制终止子的形成。当 UTP 充足,转录快速通过衰减子,mRNA 的终止子位点在核糖体结合前暴露出来,形成终止。

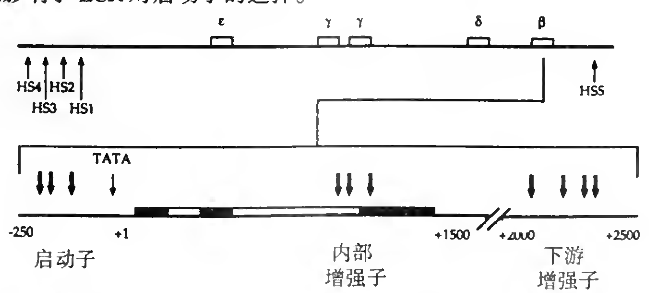
文框 29.1 真核基因调节——β珠蛋白和β珠蛋白基因簇

珠蛋白基因的组织 携带氧的血红蛋白由两个α珠蛋白和两个β珠蛋白多肽链组成。珠蛋白基因位于两个基因簇中:α珠蛋白基因簇包含了ζ-珠蛋白和α珠蛋白基因,而β珠蛋白基因簇由ε-珠蛋白、γ-珠蛋白、δ-珠蛋白和β-珠蛋白基因组成(见下面的图)。珠蛋白基因在发育中是被调节的,如在文框15.1中所讨论的。

β-珠蛋白基因的调节:启动子和增强子 β-珠蛋白基因的调节在体外缺失和基因5′端突变的实验中研究得很多。启动子跨越大约250bp,并包含TATA盒和一个CCAAT盒,及多个普遍存在的转录激活子和红细胞专一性因子GATA-1的结合位点。启动子对基本的细胞专一性转录已经足够,但高水平的转录活性依赖于两个增强子。一个部分跨越内含子2和外显子3,另一个位于基因的下游。基因间的增强子包含三个GATA-1结合位点,主要负责增加细胞专一性的表达。下游增强子包含组成性和红细胞专一性因子的结合位点,与β-珠蛋白基因表达的短暂调节有关。

β-珠蛋白基因簇基因座控制区域(LCR) 红细胞中所有β-珠蛋白基因高表达所需的基因座控制区域位于大约ε-珠蛋白基因(β-珠蛋白基因簇中位于最5′端的基因)5′端上游大约20kbp。LCR缺失突变的个体基因簇中所有基因的表达消失(εγδβ地中海贫血症),并且在由启动子和增强子控制的人β-珠蛋白在转基因小鼠表达量很低,只有在转入LCR后才强表达。LCR包含四个红细胞专一性的Dnase I *高度敏感位点*(参阅),代表了转录因子所在的没有核小体的区域,在此区域有多个正调控因子的结合位点,包括组成因子和红细胞专一性的转录激活因子GATA-1和NF-E2。LCR的活性有多种解释的模型。它被认为通过形成拓扑结构分离的染色质区域,建立了一个独立的染色质结构域,并且在LCR中发现了假定的矩阵相关区域,在转基因小鼠中可作为边界元件。LCR被认为使基因簇免受染色质位置的影响,解释了它能增加转基因小鼠中整合β-珠蛋白基因的表达,而不能增加作为附加体形式存在不受位置效应影响的转染β-珠蛋白基因的表达。

除了作为一般正调节因子的作用,LCR也控制β-珠蛋白基因簇表达的时相。进一步转基因的研究表明β-珠蛋白基因根据与LCR的距离在不同发育阶段被激活。解释距离依赖激活的模型涉及了染色质重建波的传播越过基因簇和LCR与单个启动子之间的按顺序相互作用。结合在LCR上的多个转录激活因子被认为能够与单个启动子/增强子上的转录因子作用,形成所谓的**全复合物**(holocomplex),然后激活基因的表达,可能还进一步征集了重要的激活子。在珠蛋白启动子中只有在LCR中有DNase高度敏感位点的事实支持了这种解释。单个启动子之间在与LCR相互作用中存在着竞争,如同增强子竞争(参见*亲本印迹*),结果取决于单个全复合物的相对稳定性。支持这个模型的是,在果蝇中发现了与Kruppel蛋白相关的转录因子,对γ-珠蛋白到β-珠蛋白的转换起关键作用,可能是稳定了β-珠蛋白-LCR全复合物。此外其他转录因子的水平,包括GATA-1,也影响了LCR对启动子的选择。



人β珠蛋白基因簇(上面)由五个基因组成,在染色体上的排列反映了激活的顺序。在5′端上游20kbp处有基因座控制区域,其特点是有四个DNase I高度敏感位点(HS1-4)。β珠蛋白基因(下面)有三个外显子和两个内含子,为上游启动子和两个增强子所调节。红细胞专一性的转录因子GATA-1的结合位点以粗箭头表示。在基因座控制区域中还有多个GATA-1和第二个红细胞专一性蛋白,NF-E2的结合位点。

文框 29.2 细菌基因调节——*lac* 操纵子

基本结构 *lac* 操纵子包含了乳糖代谢所需的三个结构基因(参见下面的图和表),被乳糖协同诱导。这些基因按照 *lacZYA* 的顺序排列,由 *P1* 启动子转录成多顺反子。还有第二个隐含的启动子,在体内的功能还不清楚。在 *lac* 操纵元中有四个调节元件。在 *P1* 上游大约 60bp 有一个正调节元件,是与 CAP-cAMP(代谢激活蛋白)复合物结合的**激活位点**(activator site, **AS**)。*P1* 下游 11bp 是结合 Lac 阻遏蛋白的**主要操纵基因**(major operator, O_1)。还有两个次操纵基因位点(O_2 和 O_3)位于 −82和 +401 位置,也可以结合 Lac 阻遏蛋白。Lac 阻遏蛋白由 *lacI* 编码,位于操纵子上游,是组成型表达(在它的启动子中没有调节元件)。*lacI* 基因位于邻近位置这点并不重要,其他调节因子的基因座与所调节的操纵子的结构基因距离很远(如 *gal* 和 *trp* 操纵子)。

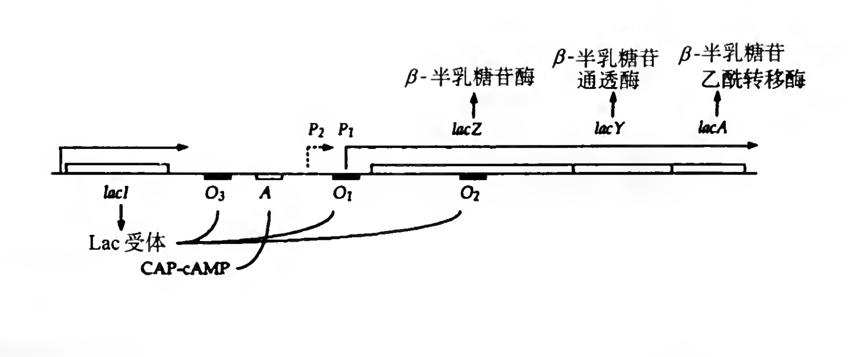
基因	酶	功能
lacZ	β-半乳糖苷酶	乳糖→β-D-半乳糖+D-葡萄糖
lacY	β-半乳糖苷通透酶	将乳糖转移到细胞内
lacA	β-半乳糖苷乙酰转移酶	不清楚,不是半乳糖代谢必需的

调节回路 由于 Lac 阻遏蛋白的合成并结合在操纵基因上抑制转录,*lac* 操纵子在正常情况下是被抑制的。乳糖的代谢产物别乳糖与阻遏蛋白结合,通过改变 DNA 结合结构域的构象而使阻遏蛋白失活。因此乳糖存在的情况下,Lac 阻遏蛋白被灭活,乳糖代谢基因去阻遏——*lac* 操纵子是受负调节的可诱导的操纵子。然而 *P1* 启动子缺乏 −35 序列,是弱启动子。为了增强转录活性,CAP-cAMP 复合物必须与激活位点(AS)结合,直接帮助 RNA 聚合酶的结合。CAP-cAMP 的活性依赖于细胞内的 cAMP 水平,后者与细胞内葡萄糖水平成反比。所以,*lac* 操纵子在乳糖存在且另一种碳源,葡萄糖缺乏(代谢抑制)时才被激活。葡萄糖如何调节 cAMP 水平目前还不清楚。

定义功能元件的突变 *lac* 操纵子中的功能元件是通过突变和**互补分析**(complementation analysis,参阅)定义的。后者需要操纵子是二倍体,在细菌中通过将操纵子作为质粒的一部分导入细胞(参见 F′ 质粒)完成。*lac* 操纵子突变可以影响单个基因或者一组基因,它们对同一细胞中的野生型等位基因可以是显性或隐性,也可以顺式或反式作用。

改变单个基因表达的突变用来对结构基因本身作图,造成功能缺失;这些突变经常是隐性的,因为同一细胞中的野生型可以提供有功能的产物。*lacZ* 和 *lacY* 的突变产生 Lac 突变表型(不能利用乳糖),而 *lacA* 突变没有表型因为这个基因并不是利用乳糖所必需的;它可能与乳糖类似物的互变有关。偶尔产生 *lacZ* 或 *lacY* 的显性突变,因为编码的蛋白以多聚体发挥功能,突变的多肽以显性负方式作用。

影响所有三个基因的突变一般定位于调节成分。启动子、激活位点和操纵基因的突变是顺式显性突变(参见)。那些影响启动子或激活位点的突变因为转录被破坏,产生 Lac 表型,而操纵基因突变由于阻遏蛋白不能结合,造成连续性表达。编码阻遏蛋白的 *lacI* 的突变以及编码 CAP 蛋白的 *crp* 基因和编码腺苷环化酶的 *cya* 突变都是反式作用。阻遏蛋白突变分成三种类型:$lacI^-$ 是隐性功能缺失突变;$lacI^{-d}$是显性失活突变——阻遏蛋白不能与操纵基因结合,将野生型阻遏蛋白阻断成为非活性的多聚体;lac^{Is}是显性不能诱导突变——阻遏蛋白与操纵基因结合但不能被诱导——这是一个**超阻遏**(superrepressor)。偶尔结构基因的突变会对操纵子中其他结构基因有顺式显性影响——这些是**极性突变**(polar mutation,参阅),它破坏下游开放阅读框架的核糖体结合位点,可能是由于突变基因二级结构的改变造成的



文框 29.3 真核生物转录因子的激活和抑制结构域

激活结构域家族 转录因子经常拥有多个不同功能的结构域——DNA 结合、转录激活、二聚化、辅因子相互作用等等。真核生物的激活结构域不是根据序列同源性，而是特定氨基酸的比例分为三个主要的家族。**酸性激活结构域**是最普遍的(单纯疱疹病毒 VP16 转录激活子，酿酒酵母 GAL4 和哺乳动物糖皮质激素受体)，包含大于 20%的谷氨酸和天冬氨酸残基。**谷氨酸富集激活结构域**(glutamine-rich activation domains)在几个同源盒和 POU 结构转录因子和结构性转录因子 Sp1 中被发现。而**脯氨酸富集激活结构域**(proline-rich activation domains)存在于 c-Jun 和 CCAAT 结合因子 CTF/NF1。人工改造蛋白的实验表明，谷氨酸富集激活结构域的活性只反映了多谷氨酸存在。相反的是，在酸性激活结构域中，除了酸性氨基酸还有很多重要的残基(转录激活的能力并不反映酸性氨基酸的数目)。

位置依赖 对不同顺式作用位点的转录因子活性研究表明三种结构域在刺激转录起始的能力是不同的。含酸性激活结构域的转录因子当与报告基因的启动子或远处增强子结合时能够激活转录。相反，含谷氨酸或脯氨酸富集结构域的转录因子当与远处的增强子结合时，不能或只有很弱的转录激活。

目标选择 含不同类型激活结构域转录因子与基本转录装置中的不同成分相互作用。组成型转录因子 Sp1 与 TFIID(TAF 的命名反映了它们相关的 RNA 聚合酶和它们的分子量，$TAF_{II}110$ 是与 RNA 聚合酶 Ⅱ 相关的 110kDal 的 TAF)中的 $TAF_{II}110$ 作用。相反几个含酸性激活结构域的转录因子主要和 $TAF_{II}40$ 作用。

抑制结构域 真核生物中与大量的激活物比较，只有少量直接转录阻遏蛋白。在已经发现的这样的蛋白中，特殊的抑制结构域经常表现出与异源蛋白的转录阻遏能力有关。几个已经被鉴定的抑制结构域，包括果蝇的 Even-skipped 蛋白，富含脯氨酸并缺乏带电氨基酸。但其他阻遏蛋白有独特的结构域，像激活结构域一样，这还有待于更多的阻遏蛋白被发现，抑制结构域的识别才会变得容易起来。

(陈　淳 译)

第30章
病毒与亚病毒感染因子

基本概念与定义

- **病毒**(viruses)是小的非细胞寄生生物。它们缺乏内源性代谢,依赖于宿主细胞提供原始材料和酶来满足基本功能——复制、转录和蛋白合成。在各类细胞中都发现了病毒的存在。感染细菌的病毒称为**细菌噬菌体**(或简称为**噬菌体**)。
- 病毒基因组是DNA或RNA,通常与蛋白质结合,形成壳[或**衣壳**(capsid)],衣壳可被脂蛋白包膜所围绕。基因组编码衣壳的结构成分和顺利完成感染周期所需的酶以及其他产物。病毒基因组的大小在1kbp到400kbp范围,最小的只有一个基因,而最大的超过300个。一个完整的病毒颗粒称为**病毒体**(virion)。
- 细胞被病毒感染通常是有害的,因为病毒常通过改变内源基因的表达模式干扰了宿主细胞的功能。在细菌中,这样的结果经常是导致生长缓慢和细胞裂解。在多细胞生物中,病毒感染经常引起局部或系统性疾病。
- 病毒可根据形态学特征(衣壳的形状,大小以及相关结构)、物理化学性质(基因组性质、病毒编码的蛋白质、脂类和碳水化合物含量)或者生物学性质(如宿主范围、转移方式、致病性)分类。有几种不同类型的衣壳:经典的二十面的衣壳(在噬菌体中可有或没有附着的尾)、球形和杆状衣壳以及螺旋(棒状)衣壳。后者的特点是它们的大小取决于病毒的核酸——衣壳是基因组周围浓缩的蛋白质亚基。其他衣壳是先形成的,然后填入核酸[**衣壳化**(encapsidation)],也就是说它们有确定的容量。有些病毒有独特的形态(如柠檬形的 *Fuselloviridae*),而有些是**多形的**(pleomorphic)或**无定形的**(amorphic)(也就是说它们缺乏结构固定的衣壳)。
- 病毒基因组的结构是多变的,反映在基因组复制和基因表达策略同样是多变的。很多病毒在它们感染循环中,逐步建立了精巧的基因表达瞬时调节的方法,避免或抵消了宿主细胞的干扰。这在真核病毒逃避免疫系统和克服细胞周期的影响及在单顺反子环境中复制和表达它们的基因中特别明显。
- 病毒在基因操作、克隆和转移中广泛被用作载体。使用最广泛的病毒包括λ噬菌体(参见分子克隆、基因组文库)、M13噬菌体(参见DNA测序、噬菌体展示)、杆状病毒、腺病毒、疱疹病毒和痘苗病毒(参见表达载体、基因治疗)和反转录病毒(参见转基因小鼠)。另外,强病毒启动子和增强子经常在质粒载体中驱动基因表达(如λ噬菌体、SV40、劳氏肉瘤病毒和花椰菜花叶病毒)。

- 很多定义为**亚病毒感染因子**(subviral agent)的物质也被描述。像病毒一样,它们是非细胞寄生物,但缺乏病毒一个或多个性质,如自身编码的衣壳,或缺乏自动感染细胞的能力。最吸引人的亚病毒感染因子是**朊病毒**(prion),它是传染性海绵状脑病的致病因素。这种生物被认为是宿主编码的神经蛋白的致病性构象异构体,能将正常异构体转变成它自己的拷贝。

30.1 病毒感染策略

病毒感染循环　　**感染循环**(infection cycle, 表 30.1)开始于病毒或其部分被导入细胞。一旦进入,病毒基因开始表达。基因表达经常组织成级联反应而病毒基因又可以分成几个组。**早期基因**(early genes)与病毒基因组的复制和**晚期基因**(late genes)的调节有关,晚期基因与子代病毒颗粒合成和基因组包装有关。研究得最清楚的例子之一的是大肠杆菌噬菌体 λ,有三组时间调节的基因(文框 30.1)。

病毒感染的结果根据病毒和宿主细胞类型不同而不同。在很多情况下,病毒感染杀死细胞——称为**裂解性感染**(lytic infection),进行这种类型感染的噬菌体称为**烈性的**(virulent)。开始感染后有一病毒繁殖的**潜伏期**(latent period),然后通过裂解从细胞中释放子代病毒。另外,病毒可以与细胞共存,通过连续的出芽或挤出产生后代——**持续感染**(persistent infection)。裂解和持续感染都是可育的,也就是说感染的子代病毒可以释放(表 30.1)。

表 30.1　增殖性病毒感染的策略(溶菌或宿存的)

感染循环的阶段	事件
附着或吸收	病毒颗粒附着到细胞外受体上
侵入	颗粒或其部分内化(至少基因组必须进入细胞)
脱壳	如果整个病毒进入细胞,从病毒体中释放基因组。真核病毒通过转移期(transport phase)将一部分脱壳病毒带到其正常复制的位置。对 DNA 病毒而言此复制处一般是在细胞核中,对 RNA 病毒而言一般是在细胞质中
早期基因表达	早期基因经常与复制和后期基因的调节有关
复制	病毒基因组产生很多拷贝
晚期基因表达	晚期基因与病毒体组装和释放有关
病毒体组装	建立新的衣壳并包装基因组
释放	子代病毒从细胞中释放

释放阶段有两个不同的结果。裂解感染引起宿主细胞被破坏,而持续感染是病毒连续从细胞中释放

很多病毒能够**潜伏感染**(latent infection):病毒在细胞中维持,但不产生有感染性的后代。在细菌中,能够潜伏感染的噬菌体被描述成温和性的,其宿主是溶原的——被称为**溶原体**(lysogen),也就是说如果重新进入裂解循环中,它能够进行裂解[这一过程称为**诱导**[1](induction)]。很多温和噬菌体通过整合到基因组中与它们的宿主溶原,整合的噬菌

1)诱导这一名称以很多方式被使用,每一意思都是"将某些打开":(1)对温和噬菌体裂解循环的诱导;(2)诱导操纵子中被阻遏基因和通过外源刺激诱导基因表达的开关(见第 29 章转录);(3)通过改变细胞在发育中的命运诱导一种细胞类型,也就是说涉及分化基因的转换(见第 6 章发育,分子机制)。

体称为**原噬菌体**(prophage)。细菌噬菌体 P1 是一个例外,它的原噬菌体是在基因组外以质粒存在。一些真核病毒能够整合,称为**前病毒**(proviruses)。然而,一个整合的病毒不一定是潜伏的。细菌噬菌体 Mu 作为原噬菌体不仅是烈性的,而且其毒性也依赖于它的整合。真核反转录病毒(参阅)也是如此。噬菌体 Mu 和真核反转录病毒两者都是可移动的因子(参阅)。动物病毒可表现出组织特异性的潜伏,如单纯疱疹病毒在神经元中是潜伏的,而在大多数其他细胞中不是这样。潜伏期可以被宿主和病毒编码的转录因子平衡的改变而破坏。裂解动物病毒在不同细胞中以不同的效率扩增。在有些情况中,根本没有感染颗粒产生〔**流产感染**(abortive infection)〕,反映了在那种细胞类型中缺乏病毒复制、基因表达或基因组包装所需的基本产物,例如在缺乏宿主编码的剪切血凝素前体的蛋白酶时,流感病毒的感染是流产的。

有些病毒在动物细胞中感染的最终结果是致瘤性转化。**转化感染**(transforming infection)是由乳多空病毒(如 SV40)、腺病毒和疱疹病毒引起的,它们都能编码使细胞周期(参阅)失去调节的产物。另外可以整合到宿主基因组中的反转录病毒也能够通过转导或激活刺激细胞生长的基因转化它们的宿主细胞。这样的基因称为癌基因(参阅)。

病毒是如何进入细胞的　　在很多宿主-病毒关系中,病毒通过与细胞表面受体结合进入细胞。在动物病毒中,这是由病毒衣壳蛋白的一个结构域介导的,称为**附着位点**(attachment site)。通常这种相互作用是**宿主范围**的唯一决定因素,而病毒的受体经常是正常细胞功能中免疫反应有关的分子(如 MHC I 类蛋白是腺病毒和 SV40 的受体,CD21 是埃-巴二氏病毒的受体,而 CD4 是 HIV 的受体)。细菌噬菌体也能吸附到表面的受体(如 λ 噬菌体附着在麦芽糖受体)上,但有些附着在其他如鞭毛或接合性伞毛这样的附器上,在后一种情况中当性伞毛被缩回细胞中时病毒被内化。因此这些噬菌体的宿主范围是由附器的存在和性质决定的〔如 f1 噬菌体被认为是雄性专一的,因为只有"雄性"细菌才有性伞毛(参见接合)〕。很多噬菌体利用**固着细胞器**(fixation organelle):它们是如尾和刺突等有利于吸附的特殊结构。

一旦与特定受体结合,病毒的内化有多种策略。在有些情况中,整个病毒并不进入细胞。很多噬菌体在细胞壁上制造一缺口,将它们的 DNA 注入细胞质中,也就是说,只有基因组进入细胞。动物病毒通过内吞被吸收。内体(endosome)的酸化使无包膜病毒脱壳,但还不清楚裸露病毒是如何在转移期中进入细胞质的。有包膜病毒与内体细胞膜融合并释放到细胞质中。

植物病毒在进入细胞质前,需要穿过不通透的外皮和植物细胞壁。正是这样的原因,病毒的感染经常是通过如昆虫这样的载体,从一个被破坏的细胞到另一细胞的转移来介导。很多植物病毒编码与胞间连丝相互作用的蛋白,使不完整的病毒体在细胞之间扩散。

30.2　复制策略的多样性

病毒基因组结构的多样性　　所有的细胞都有双链 DNA 基因组,通常在原核生物中是环状,而在真核生物是线性的,病毒在基因组性质上有更多的变化。病毒一个重要的变化是有 DNA 基因组和 RNA 基因组的区别;没有病毒是混合核酸的基因组。

每种类型的核酸都有其特定的优点和缺点,特别是对于那些感染真核细胞的病毒(表

30.2)。DNA基因组因为DNA的遗传性质(参见第16章核酸结构)和能利用宿主编码的DNA修复功能而更稳定(参见第14章诱变和DNA修复)。因此DNA基因组比RNA基因组更大而复杂,大多数复杂的DNA病毒携带了很多不必要的基因,也就是说这些基因可被宿主细胞蛋白功能所互补。还不清楚为何病毒要发展过剩的功能,尽管它们可能与病毒在不同宿主细胞环境适应的优点有关。例如,有些真核病毒编码七次跨膜结构的受体,可以改变宿主细胞的信号(见第28章信号传导)。

表30.2 DNA和RNA病毒的优点和缺点

优点	缺点
DNA病毒	
DNA是稳定的遗传物质	线性基因组的病毒需要完成5′末端复制的策略(见第26章复制)
DNA是细胞基因组的材料,因此可以利用细胞复制装置	病毒必须克服细胞周期只复制一次的限制
病毒可以利用细胞的DNA修复功能	
RNA病毒	
因为细胞RNA合成的原料是始终存在的,复制时间没有限制	RNA不如DNA稳定,没有RNA校正和修复的细胞功能。因此RNA病毒丧失功能的突变频率被限制在一定的大小范围内。很多RNA病毒因此利用空间保守的机制:重叠基因,不同的剪接,翻译移框
RNA合成是从头开始的,因此线性的基因组不需要特殊的复制策略	细胞不需要RNA复制酶,所以RNA病毒必须编码它们自身的RNA复制酶。(-)义RNA病毒必须在病毒体中携带复制酶,因为在复制酶能够合成前需要(+)义基因组的拷贝
	(只是真核)几乎所有的mRNA都是单顺反子,因为翻译在5′帽结构起始。多顺反子病毒基因组发展了克服此限制的策略,如多蛋白、片段化的基因组、亚基因组RNA的表达和内在核糖体进入位点的利用

真核病毒被真核细胞生物学属性所阻碍。DNA病毒必须克服每个细胞周期复制一次的限制,或是通过编码它们自身的复制功能,或是在宿主细胞中诱导S期。最简单的DNA病毒只能感染生长的细胞——自然生长的细胞或被病毒超感染所诱导细胞生长的。RNA病毒没有这种限制,因为它们编码自身的复制酶。然而,因为真核细胞的单顺反子环境,它们为其基因翻译发展了特殊的策略。

DNA和RNA病毒基因组都有很大的结构变化(表30.3):它们可以是环状或线性,双链或单链或部分双链。如果是单链,它们可以是**(+)义**(正义)或**(-)义**(反义)或组合的(**双义**)。基因组可以由一个染色体组成或是分隔的。在后者情况中,片段可以包装在一起或分散成颗粒[**多成分病毒**(multicomponent viruses)]。基因组可能是单一的、部分冗余或二倍体。很多病毒,例如有冗余的末端反映了它们复制或基因组包装的策略。真核RNA病毒可能在它们(+)义链加上5′帽和3′尾,两条链中有一条链加帽和加尾,或者两条链都不加。另外,5′末端可以被不同的化学基团或蛋白所阻断。

复制策略的病毒分类系统　　病毒感染的目的是基因组复制和产生很多子代病毒。通用的**病毒分类系统**(baltimore classification)在1971年制定,完全是根据复制策略(不管形态学性质和宿主范围),将病毒根据当时已知的复制策略分成五个家族。更多的发现使更多种类都需要被包括进来(表30.4)。基因组复制的策略依赖于核酸的类型、链型和正反义性。

表 30.3 病毒基因组结构——显示它们基因组的类型和构象,对真核 RNA 病毒还显示正义链和宿主 mRNA 的关系

病毒家族或种属	基因组结构	其他基因组性质
腺病毒(*Adenoviridae*)	dsDNA,线性	
乳多空病毒(*Papovaviridae*)	dsDNA,环状	
花椰菜花病毒(*Caulimoviridae*)	部分 dsDNA,环状	
多 DNA 病毒(*Polydnaviridae*)	dsDNA,环状,超螺旋	片段(很多)
细小病毒(*Parvoviridae*)	ssDNA,线性	负义
丝状病毒(*Inoviridae*)	ssDNA,环状	正义
烟草花叶病毒(*Tobamovirus*)	ssDNA,线性	正义,有帽,无 poly(A)尾
细小核糖核酸病毒(*Picomaviridae*)	ssDNA,线性	正义,无帽、poly(A)尾
双生病毒(*Geminivirdae*)	ssDNA,环状	双义,有些片段化和多成分
副粘病毒(*Paramyxoviridae*)	ssDNA,线性	负义
正粘病毒(*Orthomyxovirdae*)	ssDNA,线性	负义,片段(8)
反转录病毒(*Retroviridae*)	ssDNA,线性	正义,有帽和 poly(A)尾巴,二倍体
呼肠病毒(*Reoviridae*)	dsDNA,线性	片段(10~12)

表 30.4 病毒复制策略的病毒分类系统的延伸

类型	策略
Ⅰ,dsDNA 病毒	半不连续复制(对于细胞基因组)或者通过链置换(见第 26 章复制),如 SV40
Ⅱa,ssDNA(+)	双链复制形式的合成,从中子代 ssDNA 基因组通过链置换产生,如细菌噬菌体 M13
Ⅱb,ssDNA(-)	通过发夹引发合成双链中间体,如细小病毒
Ⅲ,dsRNA	这些病毒在病毒体中携带 RNA 复制酶,从(-)RNA 合成子代(+)RNA。(+)RNA 被包装,合成互补的(-)RNA 完成基因组,如呼肠病毒
Ⅳ,ssRNA(+)	(+)RNA 病毒有**感染性核酸**(infectious nucleic acid),因为复制酶可以直接从基因组翻译。复制酶的起始翻译后是基因组(-)RNA的产生,其作为子代(+)RNA 合成的复制中间体,如脊髓灰质炎病毒
Ⅴ,ssRNA(-)	(-)RNA 病毒有非感染性核酸,因为复制酶不能直接从 RNA 基因组中翻译。它们必须作为衣壳的一部分,携带复制酶进入细胞,产生(+)RNA,(+)RNA 同时作为 mRNA 和子代(-)RNA 合成的中间体,如流感病毒
Ⅵa,RNA 反转录病毒	(+)义 RNA 被反转录成 DNA。接着是第二链 cDNA 合成和 dsDNA 整合到宿主基因组中。原病毒的转录产生全长的子代(+)RNA,用于包装,如 HIV
Ⅵb,DNA 反转录病毒	dsDNA 转录成全长 RNA 复制中间体,再由其反转录产生一 cDNA 拷贝。第二链cDNA 合成产生子代 dsDNA 基因组用于包装。不是必须整合到宿主基因组中,如肝炎 B 病毒

双链 DNA 病毒的保留复制 大多数 dsDNA 病毒以保留方式用病毒自身或宿主编码的 DNA 聚合酶和其他复制体成分复制。环状 dsDNA 基因组病毒经常在感染早期双向复制,但在感染后期转换成滚环复制(参阅)。对如 λ 和 T7 这样的 dsDNA 细菌噬菌体,后期复制产生长的多联体,可以剪切成线性基因组大小的片段,包装到噬菌体的头部(线性基因组在感染后用包装时剪切产生的粘性末端环化)。环状 dsDNA 基因组的动物病毒——乳多空病毒(如 SV40),杆状病毒和多 DNA 病毒——主要以双向方式复制,这样环状基

因组包装进入衣壳，尽管在 SV40 感染中可以观察到滚环结构。多 DNA 病毒是独特的：它们的基因组以多个大小显著差异冗余的环状 dsDNA 出现；这些环是如何形成的还有待于研究。其他环状 dsDNA 基因组的病毒使用不同的策略：如 T4 噬菌体以多联体复制，以侵染引发（参阅）产生可利用的复制叉为中间体。

线性 dsDNA 基因组病毒可以在感染时环化（如疱疹病毒），或者可以以置换机制复制。对于染色体复制，完成 5′末端是困难的，因为末端复制不能以常规方式被引发（见第 26 章复制）。病毒不利用端粒酶。它们有不同的替代策略，包括使用末端蛋白来引发（如腺病毒），或者共价补平的发夹末端（痘病毒）。虹彩病毒（*Iridovirdae*）是感染低等脊椎动物和无脊椎动物的病毒家族，它们使用的机制涉及在细胞核中部分复制，接着转移到细胞质中复制产生大的多联体，后者加工产生基因组长度的环状未成熟和末端冗余的分子。DNA 病毒复制的核心策略见图 30.1。

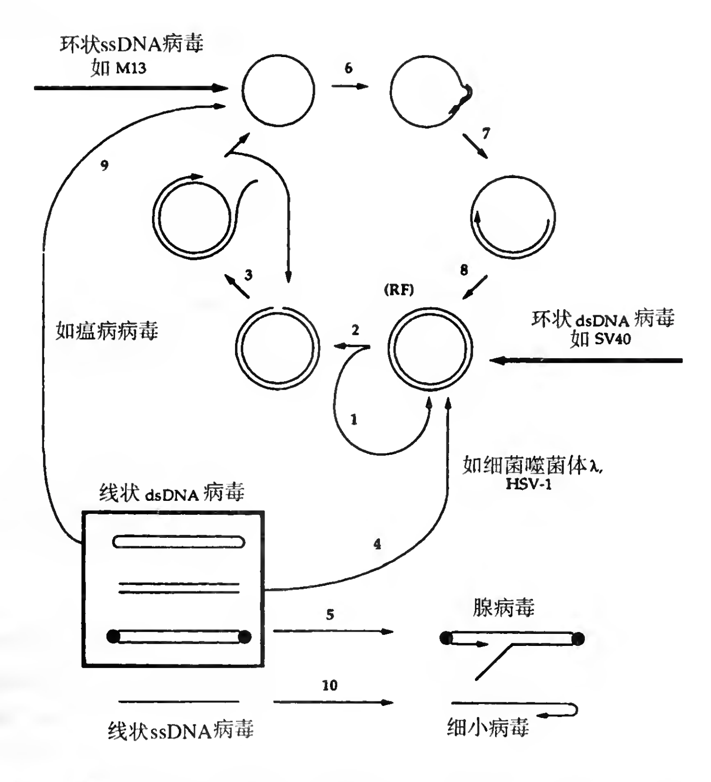


图 30.1 DNA 病毒复制策略（不包括反转录病毒）。基因组可以是双链（ds）或单链（ss），环状或线性的。环状 dsDNA 基因组病毒经常以保留双向方式复制（1），但在复制起点可以导入一缺口（2），并且它们转换成滚环复制产生多联体（3），环状 dsDNA 基因组对于某些带有线状 dsDNA 基因组的病毒而言也是一种中间体（4），尽管其他的病毒也可以使用替代策略，在末端蛋白上引发（5）。环状 ssDNA 基因组病毒也使用 dsDNA 中间体，称为**复制型**（replicative form）。一个 ssDNA 环被引发（6），以第二链合成来产生一复制型（7，8）。这样可以以常规的环状 dsDNA 基因组的复制方式产生更多的复制型（1）或通过滚环复制（2，3）产生单链子代基因组或多联体。有些线性 dsDNA 病毒有共价补平的末端，可以使用环状 ssDNA 中间体（9）。最终，线性 ssDNA 病毒可以通过发夹引发复制（10）。使用每种策略的例子在图中显示。

使用滚环复制机制的 DNA 病毒是经常产生**多联体**[1](concatemers)用于包装。多联体含有基因组的很多拷贝,以头尾连接的方式排列,可以以两种方式包装。在第一种策略中,特定的顺式作用位点可以被内切酶识别,将多联体剪切成基因组长度的片段,用于插入衣壳。这被如噬菌体 λ(*cos* 位点)使用,产生的每个病毒颗粒含有没有冗余的相同基因组。第二种策略满头装机制,可以被如 P22 噬菌体和真核蛙病毒 3 使用。它们有一顺式作用位点允许包装的起始,但一旦被识别,就没有特定有利于单个基因组包装的位点。进而,包装持续进行直到衣壳被装满,然后 DNA 被剪切。一般,衣壳的容量允许多于一个基因组被包装,这样产生**末端冗余**(terminal redundancy,在基因组末端出现了重复顺序)和**循环交替变换**(circular permutation,不同的基因组可以在不同的位置开始和结束,但含有同样的基因座)。两种包装策略在图 30.2 中比较(另见普遍性转导)。

DNA 类反转录病毒　嗜肝 DNA 病毒(如肝炎病毒 B)和花椰菜花病毒(如花椰菜花叶病毒)是 dsDNA 病毒,通过转录和反转录复制。在这两种情况中,双链 DNA 基因组是不连续的。嗜肝 DNA 病毒有环状 DNA,但不封闭,部分双链 DNA 基因组有一固定长度的反义**长链**[L(-)]和可变长度的正义**短链**[S(+)]。不同种的花椰菜花病毒基因组带有 3 个或 4 个单链不连续开放环。复制从长正义转录本的**前基因组**(pregenome)的转录开始,前基因组经常比基因组长,可同时用作 mRNA 和产生新(-)DNA 的模板,后者用病毒编码的反转录酶完成。在嗜肝 DNA 病毒中,短链合成在反转录酶相关的 RNase H 活性对前基因组降解后开始,但它的合成可以被宿主细胞的裂解所打断,结果使它有可变的长度。

转座噬菌体　细菌噬菌体 Mu 像一个转座子(参阅),只在其宿主基因组中复制,通过反复复制转座(参阅)完成;这称为**转座复制**(transpositional replication)。与其他转座子不同,Mu 编码一种功能将自身从基因组中切除,并包装到衣壳中,以噬菌体的方式感染其他细菌(见转座讨论中的可移动的遗传因子)。

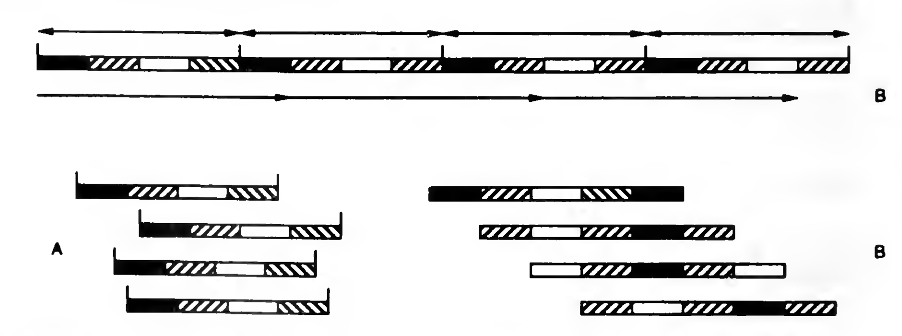


图 30.2　多联体 DNA 的包装策略。图上端的一个多联体被分成基因组长度(双向箭头),每个基因组被人为分成四部分序列块(以不同的阴影表示)。多联体的加工是通过在特定位点(A)的剪切产生非冗余和未改变的基因组用于包装。多联体的加工还通过非特异位点(B),但产生的同样大小的片段比非冗余基因组大,最后产生有末端冗余的循环交替变换的基因组。

1)**多联体**是 DNA 序列的头尾共价连接,产生多联体的加工过程称为**多联体化**。它们与连环体(catananes)不一样,连环体是通过链状排列产生的互连的 DNA 环(参见 DNA 拓扑学)。

单链 DNA 病毒的复制 单链基因组病毒通过双链中间体复制，中间体的反义链作为新DNA 基因组合成的模板(图 30.1)。细菌噬菌体属于这种类型[丝状病毒，包括丝状噬菌体 M13 和 f1 以及微小病毒(*Microviridae*)]，具有正义的单链 DNA 基因组，反基因组中间体必须用作 mRNA 合成的模板。双生病毒群(geminiviruses)是植物病毒的一个家族，有类似的基因组结构和复制机制，尽管它们的基因组是双义的，或在有些情况下，病毒是片段的和多成分的。其基因组是环状的，被称为**病毒链**或 **v-链**，而反基因组是**互补链**或 **c-链**。c-链的合成形成了双链环，称为**复制型**(replicative form, **RF**)，它可通过常规的双向机制复制产生自身的拷贝。然后，复制机制转换成不对称的滚环复制机制，产生多联体单链基因组用于包装(参见 DNA 测序)。

有些动物病毒可以类似 M13 类噬菌体的方式复制，但研究得最清楚的 ssDNA 动物病毒家族，乳多空病毒则使用一种不同的策略。乳多空病毒基因组是线性的。有些物种(如小鼠微小病毒)包装特定的(－)义 DNA；另一些物种包装两类链到不同的颗粒中，不同物种有(＋)链的病毒后代在 1%～50%之间。基因组含有自身互补的重复序列，它们通过形成发夹结构引发相反链的合成。发夹结构被病毒编码的蛋白 Rep 切除，产生基因组长度的分子用于包装。发夹结构采取特定的 T 形构象，当亲本基因组被除去后，传统的发夹在子代基因组中保留。这样就允许连续引发新的子代链，形成大的多联体。

传统单链 RNA 病毒的复制 RNA 病毒一般是单链的，经常是完全(＋)义或(－)义的(参考 ssDNA 基因组可以是双义的)。原则上(＋)义 RNA 基因组一旦进入细胞就可以作为 mRNA，而(－)义基因组却不能。因为细胞一般不编码 RNA 复制酶的功能，感染进入细胞的裸露的(＋)义 RNA 病毒基因组是感染性的，此因为可以产生病毒 RNA 复制酶，而(－)义 RNA 基因组是潜伏性的，此因在为复制酶被翻译前必须先合成基因组的互补拷贝。因此(－)义 RNA 在它们的衣壳中携带复制酶进入细胞，在病毒基因表达前合成(＋)义 RNA。大多数植物病毒是(＋)义 RNA 病毒。

(＋)义 RNA 病毒一般有线性的基因组。感染真核生物的病毒[如细小核糖核酸病毒、冠形病毒(cornavirus)，烟草花叶病毒]可能进行类似宿主 mRNAd 5′帽或 3′多聚腺苷尾巴的加工，或者也可能没有这样的结构。5′帽对细胞 mRNA 的翻译稳定性非常重要(见第 27 章 RNA 加工)，所以缺乏这种结构的病毒通常在它们基因组的 5′端以另一种方式以不同的化学基团或病毒编码的蛋白形成阻断。细小核糖核酸病毒缺乏帽结构，利用细胞对加帽 mRNA 的依赖性，通过抑制对核糖体定位关键的帽结合蛋白(参阅)来抑制宿主蛋白的合成。它们本身基因的翻译有独特的内部核糖体进入位点(见下文)来介导。大多数(＋)义 RNA 病毒用一(－)义 RNA 复制中间体复制，中间体可作为基因组复制和 mRNA 合成的模板。然而，反转录病毒用独特的涉及 dsDNA 中间体的策略来复制(见下文)。真核(－)义 RNA 病毒倾向于在它们的基因组上不加帽和多聚腺苷酸，作为 mRNA 合成的模板和产生基因组长度的(＋)义 RNA 复制中间体，中间体本身是下一步子代(－)义 RNA 基因组合成的模板。

双链 RNA 病毒的合成 几种病毒家族有双链 RNA 基因组，如感染真菌和植物的分病毒(*Rartitiviridae*)，感染单细胞真核生物的全病毒(*Totiviridae*)，囊状病毒(*Cystoviridae*)家族中的噬菌体和在多细胞生物中有广泛宿主范围并研究得最清楚的呼肠病毒

(*Reoviridae*)。所有的家族都是线性基因组,除了亚病毒(*Hypoviridae*)都是片段化的——大多数家族是两或三个片段,但在呼肠病毒中是8～10片段。真核dsRNA病毒的(+)义链倾向于加帽而不多聚腺苷酸化,而(-)义链不进行加工。呼肠病毒使用(-)链产生(+)基因组链和mRNA。因此它们是以经典的(-)义RNA病毒的方式复制,同样,它们在衣壳中携带复制酶。(+)义链包装到部分形成的衣壳中,在那儿(-)义链开始合成。

反转录病毒的复制周期 RNA反转录病毒的复制策略涉及转录、反转录、RNA降解、DNA合成和整合到宿主基因组中的有序的独特复杂加工过程(图30.3)。病毒基因组的

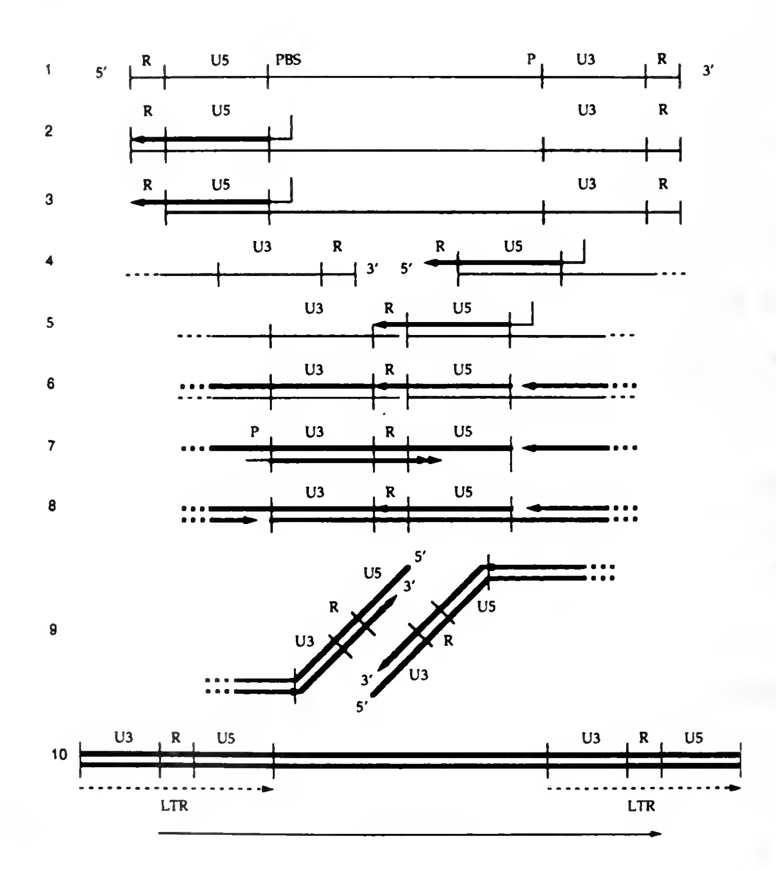


图30.3 反转录病毒的复制策略。RNA以细线、DNA以粗线表示。(1)病毒(+)RNA有冗余的末端(R)和独特的5′和3′区域(U5和U3),在5′末端加帽和3′末端多聚腺苷酸化(没有显示)。(2)(-)链DNA合成由宿主编码的tRNA引发,tRNA与U5下游的引物结合位点(PBS)结合;(-)链DNA合成进行到RNA模板的5′末端。(3)RNA模板的5′冗余区域被RNase H降解。(4)这个区域未配对DNA可以与同一基因组或另一基因组3′冗余区域配对。(5)DNA与RNA杂交允许(-)链DNA在新的RNA模板上连续合成。(6)(-)DNA链在新生模板上的延伸建立了"第一次跳跃",连续复制直到取代tRNA引物。(7)当(-)DNA链延伸时,RNA模板被RNase H降解。一些RNA片段保留,在U3上游的多嘌呤区域(P)来引发(+)DNA链的合成。(+)链DNA的合成起始,转换到起始模板,建立"第二次跳跃"。(9)通过置换复制基因组的末端区域完成(+)和(-)链DNA的合成。(10)这样产生了特征性的反转录病毒基因组长末端重复顺序(虚箭头)。反转录病毒RNA的双链cDNA拷贝能够在此位点整合到宿主基因组(参见*反转录转座*)。转录从U3中的启动子起始,在U5终止,产生病毒RNA或长箭头所示的包装,它等同于(1)中所显示的RNA链。从右侧LTR开始的转录也可以发生,可激活邻近的宿主基因(参见*慢性转化的反转录病毒*)。

中心区域 两侧是正向长末端重复序列(LTRs),它携带了负责从整合的原病毒中产生基因组长度 RNA 的调节元件。反转录病毒的末端冗余是复制策略的产物:涉及两个模板转换的拷贝选择复制。真核转座因子家族的*反转录转座子*(retrotransposons,参阅)以类似反转录病毒的方式移动,它们基因组的结构类似,区别在于常缺乏编码衣壳蛋白的 *env* 基因。反转录转座子和反转录病毒使用相似的整合机制(见第 13 章可移动的遗传因子)。人类免疫缺陷病毒(HIV)是反转录病毒,与获得性免疫缺陷症(AIDS)有关(文框 30.2)。

30.3 病毒基因表达的策略

时序控制——病毒感染的调节级联 病毒感染周期经常被分成不同的阶段,反映了特定基因亚群的表达。病毒表达的**时序控制**(temporal control)类似于发育中基因表达的控制:都涉及转录级联调节或基因表达的其他水平的调节(参见第 9 章基因表达与调节)。在 DNA 病毒中,时序控制经常是通过表达如某一个早期基因的表达产物是晚期基因的重要调节因子。较晚作用的基因可以编码进一步的调节因子,将遗传调节因子安排成相互依赖的系列,按照需要将很多时序步骤加到感染周期中。很多病毒将它们的感染周期分成两个或三个这样的步骤,不同阶段之间的延迟反映了合成或激活正确的调节因子所需的时间。

不同的病毒以不同的机制来达到这一目的。例如:λ 噬菌体、人类免疫缺陷病毒(HIV)和单纯疱疹病毒(HSV)的调节反应在文框 30.1、30.2、30.3 中分别讨论。λ 噬菌体裂解周期的进行涉及通过抗终止调节基因表达的三个时期,而裂解和溶原的选择依赖于转录因子之间的平衡。HSV 裂解感染也是由转录调节控制的,但 HIV 感染周期主要是通过调节 RNA 剪接和从细胞核中的转运控制。其他病毒使用的机制涉及 σ 因子的合成(T7 噬菌体)、翻译的调节(腺病毒)。RNA 病毒也可以使用时序调节,经常由基因组结构和是否存在复制酶调控。例如披膜病毒(togaviruses)有二顺反子(biocistronic)的基因组,但只有 5′端的基因能够在早期感染中被翻译,这是因为需要病毒复制酶(由 5′基因编码)产生较小的 RNA,它是由下游基因(编码衣壳蛋白)转录的。对双义病毒而言,早期基因产物可以从基因组链产生的 RNA 中翻译,但晚期基因的产物只能从反基因组链产生的 RNA 中翻译得到;因此复制必须在晚期基因表达之前进行。

真核 RNA 病毒的蛋白合成 真核 RNA 病毒的主要策略问题是如何在真核细胞单顺反子环境中翻译它们的开放阅读框架。细胞 mRNA 几乎都是单顺反子(见第 29 章转录),因为核糖体定位依赖于 5′修饰的帽。这与细菌中的情况不同,在细菌中核糖体可以进入内部 Shine-Dalgarno *序列*(参阅)。RNA 病毒绕开这种阻碍的策略是非常精巧的(表 30.5)。最值得注意的是在细小核糖核酸病毒中发现的情况,它有类似细菌 Shine-Dalgarno 的**内部核糖体进入位点(IRES 基序)**。这已被广泛应用于克隆基因的表达,使它们能在同一启动子控制下协同表达,这在转基因动物基因表达和报告基因的构建中有很多应用(参见*转基因小鼠*、*基因敲除*,*报告基因*)。

表 30.5 多顺反子真核 RNA 病毒中蛋白合成的策略

策略	机制和举例
片段化基因组	片段化的 RNA 基因组允许单个基因排列在分离的片段上。流感病毒有八个基因在八个反义 ssRNA片段上，编码 10 个产物(两个是通过不同的剪接产生的)。这种策略有利于通过重分配产生遗传混合。
多蛋白	整个基因组表达成一单链多肽，然后剪切产生基因产物。这种策略使所有基因产物由一个转录本编码，为细小核糖核酸病毒所使用。
IRES 基序	有些细小核糖核酸病毒基因组有 **IRES(内部核糖体进入位点)**基序，允许核糖体能够在没有反式作用成分帮助识别帽的情况下结合。IRES 应用二级结构在没有帽结合蛋白的情况下，征集翻译起始所需的因子。在重组的结构中，IRES 被置于一开放阅读框架上游，这样能在真核中表达多顺反子 mRNA。
不连续的 mRNA	弹状病毒(*Rhabdoviridae*)从它们多顺反子的反义基因组中，转录对应于每个病毒基因的单个 mRNA。转录酶连续转录每个基因，但增加了解离的可能性。这种转录延伸中新的基因调节形式使五个转录本以逐渐降低水平的方式产生。
嵌套亚基因组 RNAs	披膜病毒基因组是二顺反子的，在感染早期只有 5′端开放阅读框编码的复制酶能够被翻译。一旦复制酶被合成，它可以起始一个含有下游开放阅读框的亚基因组 mRNA 的转录，下游的开放阅读框以多蛋白编码衣壳蛋白。冠形病毒使用一类似的策略，涉及对应于七个基因的七个亚基因组 mRNA。每个都有共同的可用作引物的引导顺序。

30.4 亚病毒感染因子

亚病毒感染因子的分类　**亚病毒感染因子**是含有基因组但缺乏病毒一些重要性质的非细胞生物(表 30.6)。病毒和亚病毒感染因子区别于其他自身复制的遗传因子(如质粒和转座因子)和细胞器(如线粒体和叶绿体)的共同性质是它们以稳定的细胞外形式存在，允许它们以自己的方式进行感染。

表 30.6 病毒和不同类型的亚病毒感染因子以及它们可区分的性质

因子	有区别的性质
病毒	DNA 或 RNA 基因组 基因组在介导感染的蛋白衣壳中包装 基因组编码衣壳蛋白和感染周期所需的其他功能
类病毒(viroid)	RNA 基因组 基因组是裸露的(没有蛋白衣壳)，感染是机械性介导的 基因组似乎是不编码蛋白的
卫星病毒(satellite virus)	复制依赖于第二个不同类型的辅助病毒的共感染的病毒，辅助病毒以反式提供所缺少的功能；卫星病毒和辅助病毒是不同的实体；它们编码自身的衣壳，独立感染细胞
卫星核酸(satellite nucleicaid)	(a)有类似病毒编码性质的 DNA 或 RNA 基因组，但缺乏编码自身衣壳的能力，因此需要辅助病毒的后期功能 (b)在植物中发现的类似类病毒的 RNA，在它的辅助病毒的衣壳被附带包装，依赖于辅助病毒的感染和复制功能。辅助病毒不依赖于卫星
拟病毒(virusoid)	在植物中发现的 RNA，结构和功能类似于类病毒，可包装到病毒衣壳中，但实际上是病毒基因组的部分，不像卫星 RNA 可放弃
朊病毒(virino)	理论上的感染性颗粒，具有非编码核酸基因被包装在来源于宿主的蛋白衣壳内
朊病毒(prion)	蛋白质感染颗粒，对破坏核酸的处理耐受。一般是指整个由宿主编码蛋白组成的感染因子，这样的蛋白可以改变成致病性和自我复制的构象

类病毒:类病毒是感染植物的单链环状 RNA 因子。它们主要在三个方面不同于 RNA 病毒:大小(是最小的 RNA 病毒的四分之一,也就是说仅有 25～400 碱基);基因组不编码任何蛋白以及它们不被衣壳包装。类病毒的感染被机械性介导。

大约有 30 种不同的类病毒被发现:有些产生疾病症状,有些是潜藏的。疾病症状从温和到致死,可能是对宿主 RNA 加工的干扰所致。所有的类病毒都进行广泛的分子内碱基配对形成具有保守二级结构的硬杆状结构,据此来分辨两类类病毒家族。在混合感染中,单个类病毒同源结构域之间频繁发生的重组产生新的序列变异。

类病毒的复制需要宿主编码的 RNA 聚合酶,这样的 RNA 聚合酶有弱的 RNA 复制酶活性。类病毒被认为通过滚环复制产生多联体复制中间体,中间体作为基因组合成的模板。单链环状基因组如何从双链多联体中释放目前还不清楚,尽管鳄梨日斑病类病毒可以催化自我剪切,但是在 I 型内含子中也发现几种类病毒结构特征,表明在有些物种中自我催化单节体形成,而在其他物种中多半依赖宿主的剪切(参见核酶、自我剪接内含子)。

卫星　　卫星[1]是依赖于**辅助病毒**(helper virus)进行成功感染的亚病毒感染因子。有两种类型的卫星:**卫星病毒**编码自身的衣壳蛋白(因此它介导感染而在感染周期中有些阶段不需要帮助),**卫星核酸**不能编码。卫星病毒包括真核依赖病毒(*Dependovirus*,如腺伴随病毒)和不同辅助病毒的几种 ssRNA 植物病毒。卫星核酸可以分成两类。有些行为像病毒,但晚期功能需要辅助,因此使用辅助病毒的衣壳蛋白(如 P4 噬菌体使用 P2 噬菌体的衣壳蛋白,参见杀伤因子)。这类卫星可以被认为是质粒,这种质粒通过辅助病毒提供晚期功能转变成病毒,形成感染性颗粒——有些重组 DNA 研究所用的载体就是利用类似的原理(参见噬菌粒)。其他卫星只在植物中发现,在结构上类似类病毒,没有编码功能,在辅助病毒的衣壳中搭便车来介导感染。这些卫星的感染和复制依赖于辅助病毒,但反之不然;病毒感染有无卫星都可以发生(卫星的存在可以影响所感染植物的症状,降低辅助病毒的产量,这可能是因为在复制时对成分和原料的竞争;例如黄瓜花叶病毒对番茄的感染相对温和,但在 CMV 相关 RNA 的卫星 RNA 存在下,产生致死损伤的结果。在其他植物中,卫星 RNA 的存在降低症状的严重性)。在有些病毒中,类病毒样的因子成为基因组中的永久性成分(也就是说辅助病毒和卫星的复制和感染相互依赖)。例如鹿茸烟草花叶病毒(velvet tobacco mottle virus)有 RNA1 和 RNA2 组成的两部分基因组,后者类似类病毒,但对病毒的感染和复制是关键的;在这种情况下,RNA2 被称为**拟病毒**。相反,密切相关的苜蓿暂时性条斑病毒(lucerne transient streak)中,RNA2 不是关键的卫星。

传染性海绵状脑病因子　　**传染性海绵状脑病(TSEs)**神经系统退行性疾病,在包括人类的多种哺乳动物中自然发生(表 30.7)。TSE 因子的性质还没有完全弄清楚,但有很多证据说明它是由蛋白感染颗粒,或称为**朊病毒**(prion)引起的。朊病毒中是否存在核酸还不清楚,但朊病毒经过破坏核酸的处理,如紫外线、核酸酶作用后还具有感染性。然而,朊病毒存在不同株系,说明有遗传基础。

---

1)卫星 RNA,如苜蓿暂时性条斑病毒(lucerne transient steak virus)中的 RNA2 与卫星 DNA(参阅)没有关系,卫星 DNA 是真核基因组中高度重复的 DNA。病毒卫星不应与染色体卫星区域(参阅)混淆,卫星区域是在染色体末端发现的一种淡染的核仁形成区(参见)。

表 30.7 不同哺乳动物中的传染性海绵状脑病(TSEs)

物种	TSE	物种	TSE
绵羊和山羊	瘙痒病(scrapie)	人类	克雅氏病(CJD)
牛	牛海绵状脑病(BSE),"疯牛病"		格斯特曼综合征(GSS)
猫	猫海绵状脑病(FSE)		库鲁病(Kuru)
貂	传染性貂脑病(TME)		致死性家族性失眠征(Fatal familial insomnia, FFI)
鹿(麋鹿)	慢性损耗病(Chronic wasting disease, CWD)		

细胞编码的 **PrP^{C}**(细胞朊病毒相关蛋白,cellular prion-related protein)被认为是感染性颗粒中即使不是唯一的也是极为重要的组分。唯蛋白质假说(protein-only hypothesis)认为朊病毒中不包含核苷酸,而**非常规病毒假说**(unconventional virus hypothesis)认为朊病毒(virino)是一种异常的病毒,它的感染需要 PrP^{C}。包含上面两种假说的统一性的假说也被提出。PrP^{C} 是一种在所有的哺乳动物中都出现的神经元特异的膜缔合糖蛋白。正常的 PrP^{C} 可被蛋白酶降解,但在 TSE 病例中,神经元中出现的由 PrP 组成的纤维状积聚物(**PrP 淀粉样沉积**,PrP amyloids)具有高度的蛋白酶抗性。这种蛋白酶抗性形式被称为 **PrP^{Sc}**(瘙痒病朊病毒相关蛋白, scrapie prion-related protein)或 **PrP***,当这种蛋白从患病细胞中被分离后,发现 TSE 物质也同时被富集(虽然感染性物质与蛋白酶抗性的 PrP 的比例只有 1:10^{5})。

PrP^{C} 和 PrP^{Sc}在一级序列上是等同的,表明从正常的 PrP^{C} 转变为致病与感染性的 PrP^{Sc}是构象改变的结果。PrP^{C} 具有一个结构化的 C 端结构域和非结构化线团状的 N 端结构域,在 PrP^{Sc}中这个结构域主要呈现 β 折叠结构组织。还不清楚这种结构改变是如何发生的,但致病物质的"复制"模型认为错折叠致病形式蛋白与正常的胞内蛋白的相互作用导致被诱导再折叠,这样 PrP^{C} 转变为 PrP^{Sc}(图 30.4)。这由朊病毒疾病表现出寄主自身编码的 PrP^{Sc}的特征而非感染物质的特征这样的实验观察所支持,即内源 PrP 被转变为致病的构象。多种不同的朊病毒疾病株型的存在与非常规病毒假说一致,但唯蛋白假说可以用每种株型代表 PrP^{Sc}的不同构象形式,并可以自主催化正常的 PrP^{C} 异构体转化为这种构象的拷贝来解释。在不同种属间的感染往往出现相当长的延迟期,表明外源朊病毒转化寄主的 PrP^{C} 是一个很慢的过程。

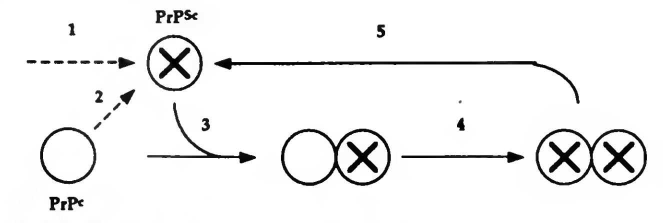


图 30.4 朊病毒复制的重折叠模型。PrP^{C} 是正常的蛋白,而 PrP^{Sc}是致病形式。起始 PrP^{Sc}可能来源于外部,即通过食物链水平传递,或在繁殖中纵向传递(1)。另外如果特定生殖细胞或体细胞的 PrP^{C} 编码基因发生突变或者其他环境因素的原因,PrP^{C} 可以自发转变为 PrP^{Sc}(2)。这种疾病通过 PrP^{Sc}与 PrP^{C} 作用的链式反应扩增发生,(3)使 PrP^{C} 转变为 PrP^{Sc}(4),这样提供了更多的 PrP^{Sc}反馈,与更多的 PrP^{C} 作用(5)。起先形成的朊病毒非常少,但会不断积累而最终指数增加。

遗传研究表明 10%的克雅氏病(CJD)病例(以及相关的格斯特曼综合征与很多病例)可以与编码 PrP^{C} 的 *PRNP* 基因的生殖细胞突变相联系。在同源的 *Prn-p* 基因座上携带突变的小鼠对瘙痒病的感染延迟时间大大缩短,表明突变可以使正常的朊病毒蛋白更易

发生由不同来源的 PrP^{Sc}诱导的构象改变。携带仓鼠编码 PrP^{C} 基因的转基因小鼠(参阅)比野生型小鼠对仓鼠感染物质更敏感。更为重要的是,*Prn-p* 基因剔除小鼠(参阅)对瘙痒病感染具有抗性,因为 TSE 物质不能复制。另外这种突变小鼠只出现温和的表型(睡眠模式改变,受损的潜伏期长)因而 PrP^{C} 内源的功能仍不清楚。有可能 PrP 编码基因的突变使 PrP^{C} 倾向于自发构象改变,从而在感染细胞中起始一扩增的链式反应。在 CJD 的自发病例中,一体细胞突变可起始这种感染,并通过越膜的蛋白-蛋白相互作用介导传播到相邻的细胞中。很明显朊病毒相关物质可以通过食物链水平地传播到人类中,如通过食用已感染的人类脑组织 (库鲁病)或感染牛海绵状脑病的牛肉(**新克雅氏病**,new variant CJD)。同时不断增加的证据也表明体细胞朊病毒疾病可以从母代向子代垂直传播。

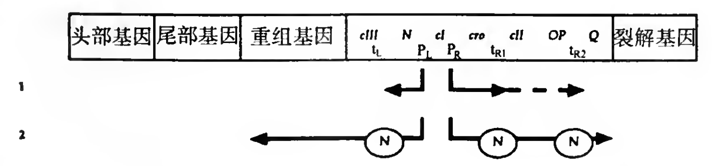
文框 30.1　λ 噬菌体

早期事件:早早期基因和晚早期基因的表达　λ 噬菌体是大肠杆菌温和噬菌体。当 λ 噬菌体感染细胞,它可以选择复制最终裂解细胞(**裂解周期**)或者整合到基因组中成为潜伏(**溶原**)状。不管 λ 噬菌体是裂解还是溶原,感染的早期事件是相同的。在进入和基因组环化后,启动子 P_L 和 P_R 被宿主 RNA 聚合酶启动。这些启动子位于 *cI* 基因两侧,转录是向外进行的(也就是说远离 *cI*),分别在位于左侧 *N* 基因和右侧 *cro* 基因边上的依赖 ρ 的转录终止位点 t_L 和 t_{R1}(参阅)结束。*N* 和 *cro* 因此称为**早早期基因**:它们在感染后立即表达。偶尔,右侧的转录通过 t_{R1} 和 t_{R2},允许编码调节蛋白的 *cII* 基因和起始 λ 噬菌体复制的 *O* 和 *P* 基因的转录。

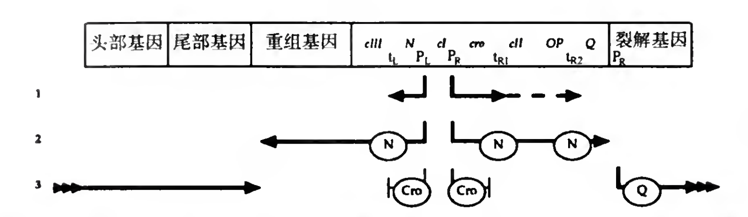
N 是抗终止蛋白(参阅),允许从 P_L 和 P_R 起始的转录越过终止位点。因此,一旦 N 被合成,从 P_L 进行的左侧转录不仅使 N 基因,而且使 *cⅢ* 和一系列涉及重组功能的基因转录,而从 P_R 进行的右侧进行的转录不仅使 *cro*,*cⅡ*,*O* 和 *P*,而且还有调节晚期基因表达的 *Q* 表达。因此 *cⅡ*,*cⅢ*,*O*,*P*,*Q* 和重组基因被称为**晚早期基因**。

裂解周期　裂解周期的特征是噬菌体复制和**晚期基因**的表达。晚期基因编码噬菌体颗粒的成分和噬菌体组建、染色体包装和宿主细胞裂解所需的蛋白质(见下图)。

复制功能由晚早期基因 *O* 和 *P* 编码,尽管还需要一系列宿主蛋白。晚期基因在单个操纵子中编码,在启动子 P_R 起始转录。转录进行到线性 λ 噬菌体右侧边缘,通过 *cos* 位点进入左臂,在不固定的读框位置终止,将尾部基因和 *att* 位点分开。在感染早期,从启动子 $P_{R'}$ 开始的右侧转录在到达终止位点只有大约 100bp。晚早期基因 *Q* 的产物是抗终止蛋白,允许对这个终止子的通读而进入晚期基因操纵子。



λ 噬菌体感染后早期基因的表达。(1)起始,从启动子 P_L 开始的左侧表达在 t_L 终止,从 P_R 开始的右侧表达在 t_{R1}(偶尔在 t_{R2})终止,使早早期基因 *N* 和 *cro* 表达。(2)N 是抗终止蛋白,允许通读终止位点,因此晚早期基因表达,包括调节基因 *cII* 和 *Q*。在 λ 噬菌体图中,基因在上排显示,而调节元件在下排。转录为粗箭头,调节因子是环。



裂解周期的遗传级联：早期事件(1)和(2)在裂解和溶原途径是相同的。这些早期事件帮助抗终止蛋白 Q 的合成以及阻遏蛋白 Cro 的转录。(3)Q 允许从 $P_{R'}$ 的转录通过附近的终止位点，因此有利于晚期基因的表达。注意晚期操纵基因进行到线性图谱的右侧，穿过左侧的头部和尾部基因，这是因为此时 λ 噬菌体是环状的。其间 Cro 与邻近启动子 P_L 和 P_R 的操纵基因结合，阻止将要建立溶原的早期基因的进一步表达。

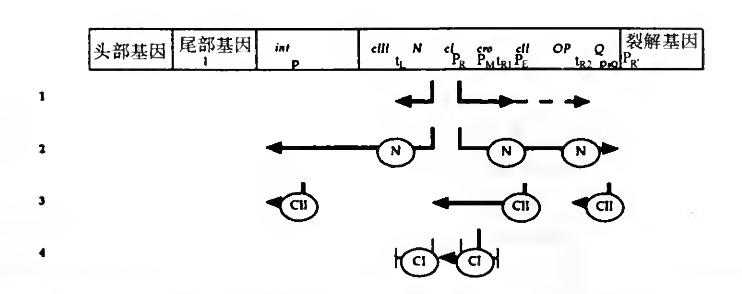
成功进入裂解周期依赖于 *cro* 的表达，*cro* 编码一种与操纵基因序列 O_L 和 O_R 结合的转录阻遏蛋白，这个操纵基因与早期启动子重叠。Cro 活性阻止调节蛋白 CI 和 CII 的表达，CI 和 CII 的功能是建立溶原，这在下文中讨论。

DNA 复制起始是双向的，在裂解阶段转换成滚环复制(见复制)。这种转换的机制还不清楚，但两种类型的复制在同一区域开始。滚环复制产生长的 λ 噬菌体基因组的多联体，然后由末端酶(由 Nu1 和 A 蛋白组成的酶)在 *cos* 位点剪切。左侧粘性末端首先被包装，然后进行头部填充直到另一个 *cos* 位点进入。裂解从细胞中释放大约 100 个子代噬菌体以及未包装的基因组和噬菌体成分。

溶原和对超感染的免疫　　溶原的特征是转录抑制和 λ 噬菌体基因组整合到细菌基因组中。这是由两个转录因子 CI 和 CII 活性的组合控制的(见下图)。

CI 是 λ 噬菌体维持潜伏状态(转录失活)的**阻遏物**。CI 与邻近早期启动基因 P_L 和 P_R 的操纵基因 O_L 和 O_R 结合，阻止了转录的进行。这样阻止了所有基因的表达，特别是 *N* 和 *Q*，保证晚期基因被阻遏。CI 与 O_R 结合的另一个结果是激活了从邻近启动子 P_M 的转录。这样有利于 *cI* 基因本身的转录。因此 CI 能够以正反馈维持自身的合成，称为**维持循环**。这可以解释对**超感染的免疫**(λ 噬菌体溶原的细菌不能再被 λ 噬菌体感染裂解)：过剩的 CI 保证了任何进入的噬菌体基因组在进入后立即被阻遏。CI 因此同时以转录激活因子和转录抑制因子发挥功能，前者是通过征集 RNA 聚合酶到启动子上，后者是通过空间阻碍。

尽管正反馈通路说明了 CI 的表达是如何维持的，但不能解释它是如何起始的。这需要转录调节因子 CII，CII 是晚早期基因的产物。CII 与三个启动子：P_E、P_{aQ}和 P_I 结合。P_E 启动子允许 *cI* 的左侧转录，通过反向转录在阻遏 *cro* 的同时建立 CI 的合成。一旦 CI 合成，它阻断了从 P_R 的转录，关闭了 CII 的合成，但在这以前 CI 的维持循环已经开始运行。P_E 启动子比 P_M 强，提供了大量阻遏物的合成，将噬菌体带到溶原状态，而 P_M 提供了低水平的持续表达以维持这一状态。P_E 启动子有很少保守顺序，这解释了为何需要 CII。P_{aQ}的转录从 Q 基因区域产生了一反义 RNA。这对任何已经合成的 Q mRNA 翻译有干扰，这样提供了第二个阻断晚期基因表达的机制。最后，从 P_I 的转录是在 *xis* 基因区域中，帮助编码将 λ 噬菌体基因组整合到宿主基因组所需的整合酶 *int* 基因表达(对 λ 噬菌体的整合机制参见位点特异重组)。



建立溶原的遗传级联:早期事件在(1)和(2)中显示裂解和溶原途径是相同的。这些早期事件有利于转录调节因子 CII 的合成。(3)CII 与三个启动子结合。在 P_I 它帮助编码将 λ 噬菌体基因组整合到宿主基因组所需的整合酶 *int* 基因表达。在 P_E 它建立 CI 的表达,通过反向转录阻遏 *cro* 的表达。在 P_{aQ} 它转录一反义 RNA 阻遏了 Q 蛋白的合成,因此阻止了晚期基因的表达。(4)CI 与邻近启动子 P_L 和 P_R 的操纵基因结合,阻止进一步的早期基因表达,但有利于自身从启动子 PM 维持 CI 的基因表达,因此建立了维持溶原的自身调节循环。

裂解和溶原之间的选择 λ 噬菌体感染后进入裂解或溶原。当 *cII* 基因表达就发生溶原。CII 通过反义抑制 Q 的表达,阻止了晚期基因的表达,帮助整合酶蛋白的合成,允许原噬菌体的插入,并建立了 *cI* 的表达。一旦 CI 被合成,它通过维持循环调节其自身的合成,通过与 O_L 和 O_R 的结合关闭了所有其他噬菌体基因的表达。当 *cro* 表达时发生裂解。Cor 通过与 O_L 和 O_R 的结合,阻断 *cI* 维持性表达,通过降低从 P_L 和 P_R 的转录阻止 *cI* 建立转录、整合酶合成和对 Q 的反义抑制,从而消除了 *cII* 的表达。

因此裂解和溶原的选择变成是 CI 或 Cro 与 O_L 和 O_R 操纵基因位点的结合,这些位点都有三个结合基序。CI 和 Cro 的不同性质和它们与操纵基因结合的不同方式决定它们如何影响转录,从而控制随之发生的事件。每个调节因子的优势又反映了 CII 是否存在:在 CII 存在时,对 *cI* 的表达有利而 *cro* 被阻遏,而在其缺乏时,*cI* 的表达不能建立,而 *cro* 被表达。*cII* 是早期基因,在感染后立即表达。然而它的产物非常不稳定,被细胞内蛋白酶快速降解。下一步的蛋白 CIII 保护 CII 免受这些影响,使它更稳定,正如宿主编码的调节因子 CAT-cAMP(参见代谢抑制)。裂解和溶原的选择因此反映了细胞中 CIII 和 CAT-cAMP 的丰度。这些因子控制了 CII 的稳定性,因此决定了 CI 是否合成。当 CIII 或 CAT-cAMP 高浓度出现时(高的感染复数(high multiplicity of infection)或不利的生长环境,也就是葡萄糖缺乏时)对溶原是有利的,而裂解在大多数其他条件下是有利的。控制 λ 噬菌体感染周期的调节步骤在下图中总结。

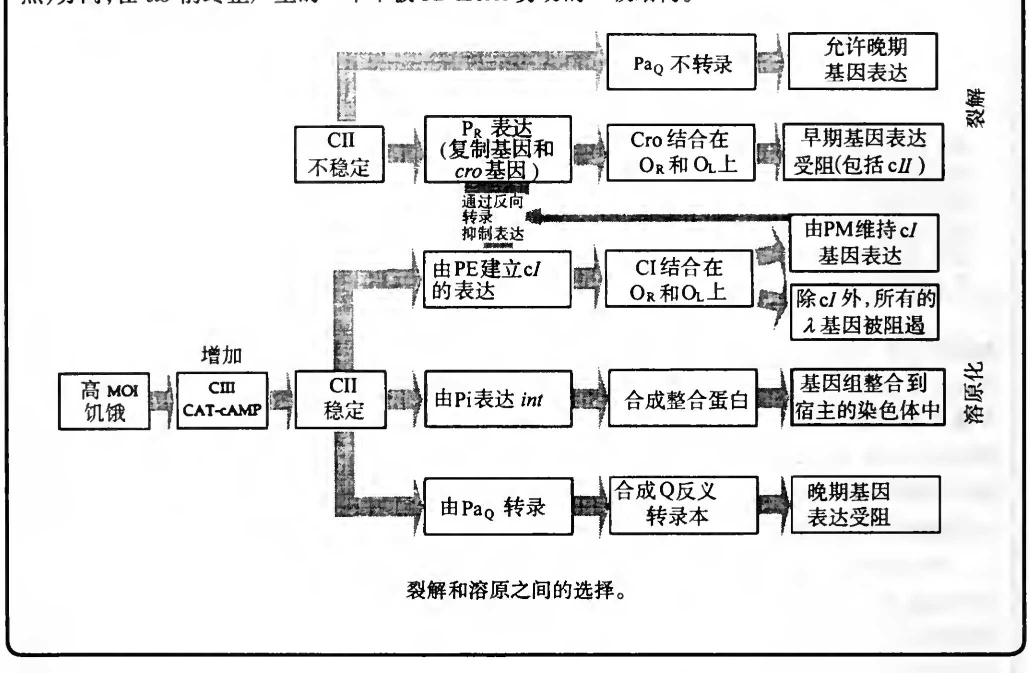
诱导 诱导是指将 λ 原噬菌体从宿主基因组中切离,进入裂解周期。诱导以低频率自发发生,可能反映了少量 CI 抑制因子从 O_R 上解离的事件,导致 Cro 的合成。诱导可以通过用破坏 DNA 的试剂处理细胞刺激发生,这是因为涉及了 SOS 反应(参阅)的激活。SOS 基因通常是通过转录阻遏物 LexA 而沉默。DNA 的损伤激活了 RecA 蛋白(参阅),RecA 蛋白剪切 LexA,诱导了 SOS 基因。CI 阻遏蛋白与 LexA 有些结构性质相似,也被激活的 RecA 剪切。这引起了阻遏物从操纵基因位点上解离,关闭了从 P_M 上的 *cI* 转录,激活了 P_L 和 P_R 启动子的转录,所以进入裂解周期。

整合和切离的控制 在溶原有利时,λ 噬菌体整合到宿主染色体中,当情况有利于诱导时,发生切离。这两个过程都需要 *int* 基因编码的整合酶蛋白,而只有切离需要外切酶,切离酶是 *xis* 基因产物。整合和切离的控制是转录和转录后控制的。

在裂解周期中,从 P_L 的转录产生编码 *int* 和 *xis* 的转录本。然而因为转录本在称为 sib 的区域终止,使 mRNA 形成一种不同的二级结构使蛋白不能被合成。mRNA 被宿主 RNaseIII 剪切,转录本包括 *int* 和 *xis* 基因座的末梢部分被降解,这样的调节机制称为反向调节(参阅)。

当对溶原有利时，CII 从 *xis* 基因中的 P_I 启动子转录激活 *int*。因此 *int* 表达，但 *xis* 不表达，只有整合酶被合成。

在诱导时，从 P_L 的转录去阻遏，但这时整合酶和切离酶都合成，允许原噬菌体切离。这时 *int* 和 *xis* 基因不进行反向调节，这是因为整合将这两个基因与 *sib*(位于 *int/xis* 和 *sib* 之间的 *att* 位点)分离，在 *sib* 前终止产生的一个不被 RNaseIII 剪切的二级结构。



裂解和溶原之间的选择。

文框 30.2 HIV 与 AIDS 的分子生物学

HIV 的感染周期 人类免疫缺陷病毒(HIV-1 和 HIV-2)是和慢病毒类(lentivirus)中的反转录病毒紧密相关的，慢病毒类也包括了猿免疫缺陷病毒(SIV)。一般的反转录病毒复制策略在正文中已经讨论，HIV 也进行同样的周期。然而，慢病毒类有比其他反转录病毒更为复杂的基因组结构，除了 *gag*、*pol* 和 *env* 基因外，编码了很多调节基因产物。这使它们能够在潜伏和裂解感染策略中转换。

HIV 是感染表面具有 CD4 受体的 T 细胞的嗜淋巴细胞病毒。HIV 的感染经常引起短期的相对温和症状的疾病。然而，CD4 细胞在 HIV 感染中被消耗，这是因为它们被裂解感染和成为免疫反应的靶。因为 T 细胞是调节免疫反应的细胞，所以 HIV 感染的第二个结果或长期结果是**获得性免疫缺陷症(AIDS)**。HIV 感染也与神经疾病有关，这些疾病可能是 HIV 感染携带 CD4 的神经元或病毒通过其他受体与脑细胞作用的结果。反转录病毒复制周期使病毒在宿主基因组中休眠很长时期，与产生 CD4 的 T 细胞结合使发展有效治疗 HIV 感染的方法非常困难。另外，病毒编码的反转录酶有很高的差错倾向，使在特定个体中感染的病毒群内产生了遗传多样性，可逃避免疫系统——正是这个原因，很难开发高效疫苗。

HIV 基因表达和调节 HIV 病毒基因组含有很多在其他反转录病毒中没有的基因。HIV-1 和 HIV-2 都携带 *tat*，*rev*，*nef*，*vpr* 和 *vif* 基因；另外 HIV-1 有独特的 *vpu* 基因而 HIV-2 有 *vpx*

基因。基因 *tat*，*rev* 和 *nef* 编码在 HIV 感染周期中有调节功能的蛋白。Tat 蛋白对复制和顺式元件的结合是必要的，*Tar* 在长末端重复顺序中发现。*Tar* 形成一复杂的二级结构抑制转录和蛋白质的合成，后者可能是通过定向 RNA 使其降解来完成的。Tat 与原病毒 DNA 及其转录本结合，减弱抑制效果。

Rev 蛋白也是转录后调节因子，通过控制细胞核中剪接中间体的转运起作用。正常情况下，在细胞 mRNA 前体剪接处有剪接成分的出现是不能运出细胞核进行翻译的。在 HIV 感染中，完全剪接的转录本只编码调节蛋白，因此没有 Rev 蛋白的情况下，就没有结构蛋白产生。Rev 与 Rev 反应元件，**抗抑制顺序**结合，允许未剪接和部分剪接的 RNA 外运，因此合成了结构蛋白。

Nef 是负作用的调节因子，可以阻止 HIV 生长失控，生长失控会引起宿主免疫反应增强。Vif、Vpu 和 Vpr 通过调节病毒基因表达或成熟病毒体的组建刺激有感染性的子代病毒产生。它们的精确作用机制还不清楚。

裂解还是潜伏　与噬菌体 λ 进入裂解和溶原依赖于环境一样，HIV 也能够对外在信号作出反应。HIV 原噬菌体的转录导致成熟子代病毒体的产生，但也增加了 Rev 的水平，Rev 通过刺激编码结构蛋白的未剪接转录本的外运，关闭了 Rev 和 Tat 蛋白的合成。Tat 为转录和蛋白质合成所需，因此降低 Tat 水平关闭了病毒的表达，造成潜伏。病毒可以对识别 LTR 启动子的宿主编码的转录调节因子反应，再次进入裂解。这种信号传导途径被认为是通过抗原与 T 细胞受体结合起始的。因此 HIV 病毒在转录被抑制前，从免疫激活 T 细胞中短暂大量释放，然后病毒再次潜伏。可能是这些短暂爆发以及病毒的遗传变异使它逃避宿主免疫反应。

文框 30.3　单纯疱疹病毒 **1**(HSV 1)的感染周期

裂解周期　单纯疱疹病毒 1 是 α 型疱疹病毒，有 150kbp 双链线性 DNA 基因组。它感染很多物种的不同类型的细胞，这是因为它与细胞表面的硫酸乙酰肝素相互作用。吸收涉及与 FGF 受体的相互作用。

一旦它进入细胞，线性基因组从核壳释放出进入细胞核，在那儿它立即被环化。核壳含有两个蛋白：一个称为**病毒体宿主关闭蛋白**(virion host shut-off protein，VHS)和 **VP16**，VP16 与宿主编码的 Oct-1 转录因子协同作用于转录 HSV-1 基因组的 5 个**早早期基因**或 **α 基因**。大多数 α 基因的产物是遗传调节因子，它们本身或是下游的 **β 基因**与 DNA 的代谢和复制功能有关。一旦这些阶段开始，一大套约 40 个**晚期基因**(**γ 基因**)被激活，产生涉及 DNA 包装和病毒体组建的蛋白。子代病毒体通过内质网转运到细胞表面。

潜伏感染　在大多数细胞中发生的是 HSV-1 裂解感染，但在神经元中，感染是潜伏的。细胞特异潜伏的基础是**潜伏相关转录本**(LATs)的转录和其后的剪接。LAT 转录单位与早早期基因 α0 重叠，但在相反方向。LATs 在裂解感染中产生，但不剪接，并运输到细胞质中。在潜伏感染时，剪接的 LATs 限制在细胞核中，可能以某种方式下调裂解基因的表达。但没有发现 LAT 相关多肽，所以 LATs 的功能可能是在 RNA 水平上的。LATs 不是潜伏感染建立所需的，但对裂解周期是关键的。在潜伏感染中，HSV-1 维持在基因组外，不复制。因此这种病毒是在活体动物中基因转移到神经元的有效载体，被用于*基因治疗*(参阅)。重组的 HSV-1 载体可通过将外源基因插入到 LAT 启动子中来进行构建。

(陈　淳 译)

参考文献和推荐读物

第 1 章　生物学的遗传和变异
Biological Heredity and Variation

参考文献

Connor, J.M. and Ferguson-Smith, M.A. (1994) *Essential Medical Genetics*. 4th edn. Blackwell Science, Oxford.

Falconer, D.S. and Mackay T.F.C. (1996) *Introduction to Quantitative Genetics*. 4th edn. Longman Group, Harlow.

Fincham, J.R.S. (1994) *Genetic Analysis*. Blackwell Science, Oxford.

McKusick, V.A. (ed.) (1996) *Mendelian Inheritance in Man*. 12th edn. Johns Hopkins University Press, Baltimore, MD.

推荐读物

Avery, L. and Wasserman, S. (1992) Ordering gene function — the interpretation of epistasis in regulatory hierarchies. *Trends Genet*. **8**: 312–316.

Frankel, W.N. (1995) Taking stock of complex trait genetics in mice. *Trends Genet*. **11**: 471–477.

Guarente, L. (1993) Synthetic enhancement in gene activation — a genetic tool come of age. *Trends Genet*. **9**: 362–366.

Hodgkin, J. (1993) Fluxes, doses and poisons — molecular perspectives on dominance. *Trends Genet*. **9**: 1–2.

Lyttle, T.W. (1993) Cheaters sometimes prosper — distortion of Mendelian segregation by meiotic drive. *Trends Genet*. **9**: 205–210.

Mackay, T.F.C. (1995) The genetic basis of quantitative variation — numbers of sensory bristles of *Drosophila melanogaster* as a model system. *Trends Genet*. **11**: 464–470.

Weeks, D.E. and Lathrop, G.M. (1996) Polygenic disease — methods for mapping complex disease traits. *Trends Genet*. **11**: 513–519.

Wilkie, A.O.M. (1994) The molecular basis of genetic dominance. *J. Med. Genet*. **31**: 89–98.

Wolf, U. (1995) The genetic contribution to phenotype. *Hum. Genet*. **95**: 127–148.

第 2 章　细胞周期
The Cell Cycle

参考文献

Hartwell, L. (1995) Introduction to cell cycle controls. In: *Cell Cycle Control: Frontiers in Molecular Biology* (eds C.M. Hutchinson and D.M. Glover), pp. 1–15. IRL Press, Oxford.

Murray, A.W. and Hunt, T. (1993) *The Cell Cycle: An Introduction*. W.H. Freeman, New York.

Osmani, S.A. and Xiang, S.Y. (1997) Targets of checkpoints controlling mitosis: Lessons from lower eukaryotes. *Trends Cell Biol*. **7**: 283–288.

推荐读物

Baylin, S.B. (1997) Tying it all together: Epigenetics, genetics, cell cycle and cancer. *Science* **277**: 1948–1949.

Donachie, W.D. (1993) The cell cycle of *E. coli*. *Annu. Rev. Microbiol*. **47**: 199–230.

Dynlacht, B.D. (1997) Regulation of transcription by proteins that control the cell cycle. *Nature* **389**: 149–152.

Gottesfeld, J.M. and Forbes, D.J. (1997) Mitotic repression of the transcriptional machinery. *Trends Biochem. Sci*. **22**: 197–202.

Lane, H.A. and Nigg, E.A. (1997) Cell cycle control: POLO-like kinases join the outer circle. *Trends Cell Biol*. **7**: 63–68.

Lutkenhaus, J. (1997) Bacterial cytokinesis: Let the light shine in. *Curr. Biol*. **7**: R573–R575.

McCall, K. and Steller, H. (1997) Facing death in the fly: Genetic analysis of apoptosis in *Drosophila*. *Trends Genet*. 13: 222–226.

Nasmyth, K. (1996) At the heart of the budding yeast cell cycle. *Trends Genet*. **12**: 405–412.

Noble, M.E.M., Endicott, J.A., Brown, N.R. and Johnson, L.N. (1997) The cyclin box fold: protein recognition in the cell cycle and transcriptional control. *Trends Biochem. Sci*. **22**: 482–487.

Pines, J. and Hunter, T. (1995) Cyclin-dependent kinases: An embarrassment of riches? In: *Cell Cycle Control: Frontiers in Molecular Biology* (eds C.M. Hutchinson and D.M. Glover), pp. 144–176. IRL Press, Oxford.

Rothfield, L.I. and Garcia-Lara, J. (1996) Cell division. In: *Regulation of Gene Expression in Escherichia coli* (eds E.C.C. Lin and A.S. Lynch), pp. 547–570. R. G. Landes/Chapman & Hall, New York.

Sherr, C.J. and Roberts, J.M. (1995) Inhibitors of mammalian cyclin dependent kinases. *Genes Devel.* **9**: 1144–1163.

Vaux, D.L. and Strasser, A. (1996) The molecular biology of apoptosis. *Proc. Natl Acad. Sci. USA.* **93**: 2239–2244.

Wake, R.G. and Errington, J. (1995) Chromosome partitioning in bacteria. *Annu. Rev. Genet.* **29**: 41–67.

第3章 染色质
Chromatin

参考文献

Elgin, S.C.R. (Ed.) (1995) *Chromatin Structure and Gene Expression*. IRL Press, Oxford.

Van Holde, K.E. (1989) *Chromatin*. Springer, New York.

Wolffe, A.P. (1992) *Chromatin: Structure and Function*. Academic Press, London.

推荐读物

Geyer, P.K. (1997) The role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.* **7**: 242–248.

Hartzog, G.A. and Winston, F. (1997) Nucleosomes and transcription: Recent lessons from genetics. *Curr. Opin. Genet. Dev.* **7**: 192–198.

Loo, S. and Rine, J. (1995) Silencing and heritable domains of gene expression. *Annu. Rev. Cell Biol.* **11**: 519–548.

Nash, H.A. (1996) The HU and IHF proteins: Accessory factors for complex protein–DNA assemblies. In: *Regulation of Gene Expression in Escherichia coli* (eds E.C.C. Lin and A.S. Lynch), pp. 149–180. R.G. Landes/Chapman and Hall, New York.

Pirrotta, V. (1997) Chromatin-silencing mechanisms in *Drosophila* maintain patterns of gene expression. *Trends Genet.* **13**: 314–318.

Sherman, J.M. and Pillus, L. (1997) An uncertain silence. *Trends Genet.* **13**: 308–313.

Tsukiyama, T. and Wu, C. (1997) Chromatin remodeling and transcription. *Curr. Opin. Genet. Dev.* **7**: 182–191.

Wade, P.A., Pruss, D. and Wolffe, A.P. (1997) Histone acetylation: Chromatin in action. *Trends Biochem. Sci.* **22**: 128–132.

Widom, J. (1997) Chromatin: The nucleosome unwrapped. *Curr. Biol.* **7**: R653–655.

Wolffe, A.P. and Pruss, D. (1996) Deviant nucleosomes — the functional specialization of chromatin. *Trends Genet.* **12**: 58–62.

第4章 染色体突变
Chromosome Mutation

参考文献

Borgankar, D.S. (1989) *Chromosome Variation in Man: A Catalogue of Chromosome Variants and Anomolies*. Liss, New York.

Therman, E. and Susman, M. (1993) *Human Chromosomes. Structure, Behaviour and Effects.* 3rd edn. Springer, Berlin.

推荐读物

Antonarakis, S.E. (1993) Human chromosome 21 — genome mapping and exploration, circa 1993. *Trends Genet.* **9**: 142–148.

Epstein, C.J. (1988). Mechanisms of the effects of aneuploidy in mammals. *Annu. Rev. Genet.* **22**: 51–75.

Hassold, T.J. and Jacobs, P.A. (1984) Trisomy in man. *Annu. Rev. Genet.* **18**: 69–98.

Moore, G., Foote, T., Helentjaris, T., Devos, K., Kurata, N. and Gale, M. (1996) Was there an ancestral cereal chromosome? *Trends Genet.* **11**: 82–83.

Zinn, A.R., Page, D.C. and Fisher, E.M.C. (1993) Turner syndrome — the case of the missing sex-chromosome. *Trends Genet.* **9**: 90–93.

第5章 染色体结构与功能
Chromosome Structure and Function

参考文献

ISCN (1995) *An International System for Human Genetic Nomenclature* (ed F. Mittelman). Karger, Basel.

Therman, E. and Susman, M. (1993) *Human Chromosomes. Structure, Behaviour and Effects.* 3rd Edn. *Springer, Berlin.*

推荐读物

Bickmore, W.A. and Sumner, A.T. (1989) Mammalian chromosome banding. *Trends Genet.* 5: 144–178.

Clarke, L. (1990) Centromeres of budding and fission yeasts. *Trends Genet.* 6: 150–154.

Greider, C.W. (1996) Telomere length regulation. *Annu. Rev. Biochem.* 65: 337–365.

Mason, J.M. and Biessmann, H. (1995) The unusual telomeres of *Drosophila*. *Trends Genet.* **11**: 58–62.

Smith, S. and de Lange, T. (1997) TRF1, a mammalian telomeric protein. *Trends Genet.* **13**: 21–26.

Zakian, V.A. (1996) Structure, function and replication of *Saccharomyces cerevisiae* telomeres. *Annu. Rev. Genet.* **30**: 141–172.

第 6 章　发育,分子机制
Development, Molecular Aspects

参考文献

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994) In: *Molecular Biology of the Cell*. 3rd edn, pp. 1037–1119. Garland Publishing, New York.

Bard, J.B.L. (ed.) (1994) *Embryos. Colour Atlas of Development*. Wolfe, London.

Gilbert, S.F. (1997) *Developmental Biology*. 5th edn. Sinauer, Sunderland MA.

Slack, J.M.W. (1991) *From Egg to Embryo: Regional Specification in Early Development*. 2nd edn. Cambridge University Press, Cambridge.

推荐读物

Cohn, M.J. and Tickle, C. (1996) Limbs — a model for pattern formation within the vertebrate body plan. *Trends Genet.* **12**: 253–257.

Coyne, R.S., Chalker, D.L. and Yao, M.C. (1996) Genome downsizing during ciliate development: Nuclear division of labour through chromosome restructuring. *Annu. Rev. Genet.* **30**: 557–578.

Dickson, B. (1995) Nuclear factors in sevenless signaling. *Trends Genet.* **11**: 106–111.

Hammerschmidt, M., Brooke, A. and McMahon, A.P. (1997) The world according to Hedgehog. *Trends Genet.* **13**: 14–21.

Joyner, A.L. (1996) *Engrailed*, *Wnt* and *Pax* genes regulate midbrain–hindbrain development. *Trends Genet.* **12**: 15–20.

Kornfield, K. (1997) Vulval development in *Caernorhabditis elegans*. *Trends Genet.* **13**: 55–61.

Leamire, P. and Kodjabachian, L. (1997) The vertebrate organizer: Structure and molecules. *Trends Genet.* **12**: 525–531.

MacOnochie, M., Nonchev, S., Morrison, A. and Krumlauf, R. (1996) Paralogous *Hox* genes: Function and regulation. *Annu. Rev. Genet.* **30**: 529–556.

Moon, R.T., Brown, J.D. and Torres, M. (1997) WNTs modulate cell fate and behaviour during vertebrate development. *Trends Genet.* **13**: 157–162.

Morisato, D. and Anderson, K.V. (1995) Signaling pathways that establish the dorsal–ventral patterns of the *Drosophila* embryo. *Annu. Rev. Genet.* **29**: 371–399.

Parent, C.A. and Devreotes, P.N. (1996) Molecular genetics of signal transduction in *Dictyostelium*. *Annu. Rev. Biochem.* **65**: 411–440.

Riverapomar, R. and Jackle, H. (1996) From gradients to stripes in *Drosophila* embryogenesis — filling in the gaps. *Trends Genet.* **12**: 478–483.

Stragier, P. and Losick, R. (1996) Molecular genetics of sporulation in *Bacillus subtilis*. *Annu. Rev. Genet.* **30**: 297–341.

Werner, M.H., Huth, J.R., Gronenborn, A.M. and Clore, G.M. (1996) Molecular determinants of mammalian sex. *Trends Biochem. Sci.* **21**: 302–308.

第 7 章　DNA 甲基化及外遗传调节
DNA Methylation and Epigenetic Regulation

推荐读物

Bird, A. (1995) Gene number, noise reduction and biological complexity. *Trends Genet.* **11**: 94–100.

Hollick, J.B., Dorweiler, J.E. and Chandler, V.L. (1997) Paramutation and related allelic interactions. *Trends Genet.* **13**: 302–308.

Jaenisch, R. (1997) DNA methylation and imprinting: Why bother? *Trends Genet.* **13**: 323–329.

Lalande, M. (1996) Parental imprinting and human disease. *Annu. Rev. Genet.* **30**: 173–195.

Lee, J.T. and Jaenisch, R. (1997) The (epi)genetic control of mammalian X-chromosome inactivation. *Curr. Opin. Genet. Dev.* **7**: 274–280.

Reik, W. and Maher, E.R. (1997) Imprinting in clusters: Lessons from Beckwith–Wiedemann syndrome. *Trends Genet.* **13**: 330–334.

Richards, E.J. (1997) DNA methylation and plant development. *Trends Genet.* **13**: 319–323.

Siegfried, Z. and Cedar, H. (1997) DNA methylation: A molecular lock. *Curr. Biol.* **7**: R305–R307.

Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**: 335–340.

第 8 章　基因
The Gene

参考文献

Lewin, B. (1997). *Genes VI*. Oxford University Press, Oxford.

Singer, M. and Berg, P. (1991) *Genes and Genomes: A Changing Perspective*. University Science Books, CA.

推荐读物

Blumenthal, T. (1995) *Trans*-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet*. **11**: 132–136.

Kable, M.L., Heidmann, S. and Stuart, K.D. (1997) RNA editing: Getting U into RNA. *Trends Biochem. Sci*. **22**: 162–166.

Stewart, A. (ed) (1995) The Trends in Genetics. Genetic Nomenclature Guide. *Trends Genet*. **11** (Suppl).

第 9 章　基因表达和调节
Gene Expression and Regul ation

推荐读物

See the other chapters concerning gene regulation in this book: Transcription, RNA Processing, Protein Synthesis, Chromatin, DNA Methylation and Epigenetic Regulation, Nucleic Acid Structure, Recombination, Signal Transduction, The Cell Cycle, Oncogenes and Cancer, Development: Molecular Aspects.

Latchman, D. (1995) *Gene Regulation. A Eukaryotic Perspective*. Chapman & Hall, London.

Lewin, B. (1997) *Genes VI*. Oxford University Press, Oxford.

Lin, E.C.C. and Lynch, A.S. (Eds) (1996) *Regulation of Gene Expression* in *Escherichia coli*. Chapman & Hall, London.

第 10 章　细菌中的基因转移
Gene Transfer in Bacteria

推荐读物

Frost, L.S., Ippen-Ihler, K.A. and Skurray, R.A. (1995) Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev*. **58**: 162–210.

Lanka, E. and Wilkins, B.M. (1996) DNA processing reactions in bacterial conjugation. *Annu. Rev. Biochem*. **64**: 141–169.

Masters, M. (1985). Generalised transduction. In: *Genetics of Bacteria*, (eds Scaife, J., Leach, D. and Galizzi, A.) pp. 197–215. Academic Press, London.

Solomon, J.M. and Grossman, A.D. (1996) Who's competent and when — regulation of natural genetic competence in bacteria. *Trends Genet*. **12**: 150–155.

Summers, D.K. (1996) *Plasmid Biology*. Blackwell Science, Oxford.

第 11 章　遗传密码
The Genetic Code

推荐读物

Crick, F.C. (1990) *What Mad Pursuit: A Personal View of Scientific Discovery*. Penguin, London.

Fox, T.D. (1987) Natural variation in the genetic code. *Annu. Rev. Genet*. **21**: 67–91.

Hatfield, D. and Diamond, A. (1993) UGA: A split personality in the universal genetic code. *Trends Genet*. **9**: 69–70.

Low, S.C. and Berry, J.M. (1996) Knowing when not to stop — selenocysteine incorporation in eukaryotes. *Trends Biochem. Sci*. **21**: 203–208.

Moras, D. (1992) Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem. Sci*. **17**: 159–164.

第 12 章　基因组与作图
Genomes and Mapping

参考文献

Brown, T.A. (Ed.) (1991) *Molecular Biology Labfax*. BIOS, Oxford.

Primrose, S.B. (1995) *Principles of Genome Analysis*. Blackwell Science, Oxford.

Singer, M. and Berg, P. (1991) *Genes and Genomes*. Blackwell Science, Oxford.

推荐读物

Charlesworth, B., Sniegowski, P. and Stephan, W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215–220.

Cooke, J., Nowak, M.A., Boerlijst, M., and Maynard-Smith, J. (1997) Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends Genet*. **13**: 360–363.

Desouza, S.J., Long, M.Y. and Gilbert, W. (1996) Introns and gene evolution. *Genes to Cells* **1**: 493–505.

Dujon, B. (1996) The yeast genome project — what did we learn? *Trends Genet*. **12**: 263–270.

Elgar, G., Sandford, R., Aparicio, S., MacCrae, A., Venkatesh, B. and Brenner, S. (1996) Small is beautiful — comparative genomics with the puffer fish (*Fugu rubripes*). *Trends Genet*. **12**: 145–150.

Fickett, J.W. (1996) Finding genes by computer — the state-of-the-art. *Trends Genet*. **12**: 316–320.

Gabor Miklos, G.L. and Rubin, G.M. (1996) The role of the genome projects in determining gene function: Insights from model organisms. *Cell* **86**: 521–529.

Gardiner, K. (1996) Base composition and gene distribution: Critical patterns in mammalian genome organisation. *Trends Genet*. **12**: 519–524.

Gerhold, D. and Caskey, C.T. (1996) It's the genes! EST access to human genome content. *BioEssays* **18**: 973–981.

Goeffeau, A., Barrell, B.G., Bussey, R.W., Davis, R.W., Dujon, B., Feldmann, H., Gailbert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. (1996) Life with 6000 genes. *Science* **274**: 546–567.

Haley, C.S. (1995) Livestock QTLs — bringing home the bacon? *Trends Genet*. **11**: 488–492.

Lander, E.S. (1996) The new genomics: Global views of biology. *Science* **274**: 536–539.

O'Brien, S.J., Weinberg, J. and Lyons, L.A. (1997) Comparative genomics: Lessons from cats. *Trends Genet*. **13**: 393–399.

Okubo, K. and Matsubara, K. (1997) Complementary DNA sequence (EST) collections and the expression of information of the human genome. *FEBS Letts*. **403**: 225–229.

Postlethwait, J.H. and Talbot, W.S. (1997) Zebrafish genomics: From mutants to genes. *Trends Genet*. **13**: 183–190.

Rowen, L., Mahairas, G. and Hood, L. (1997) Sequencing the human genome. *Science* **278**: 605–607.

Stuber, C.W. (1995) Mapping and manipulating quantitative traits in maize. *Trends Gent*. **11**: 477–481.

Tang, C.M., Hood, D.W. and Moxon, E.R. (1987) *Haemophilus* influence: The impact of whole genome sequencing on microbiology. *Trends Genet*. **13**: 399–404.

第 13 章　可移动的遗传因子
Moblie Genetic Elements

参考文献

Berg, D.E. and Howe, M. (1989) *Mobile DNA*. American Society of Microbiology, Washington, DC.

Saedler, H. and Gierl, A. (1996) Transposable elements. *Curr. Top. Microbiol. Immunol*. Vol. **204**.

推荐读物

Azpiroz-Leehan, R. and Feldmann, K.A. (1997) T-DNA insertion mutagenesis in *Arabidopsis*: Going back and forth. *Trends Genet*. **13**: 335–340.

Craig, N.L. (1997) Target site selection in transposition. *Annu. Rev. Biochem*. **66**: 437–474.

Cummings, M.P. (1994) Transmission patterns of eukaryotic transposable elements — arguments for and against horizontal transfer. *Trends Ecol. Evol*. **9**: 141–145.

Finnegan, D.J. (1990) Transposable elements and DNA transposition in eukaryotes. *Curr. Opinion Cell Biol*. **2**: 475–480.

Gierl, A., Saedler, H. and Peterson, P.A. (1989) Maize transposable elements. *Annu. Rev. Genet*. **23**: 71–85.

Goff, S.P. (1992) Genetics of retroviral integration. *Annu. Rev. Genet*. **26**: 527–544.

Kleckner, N. (1990) Regulation of transposition in bacteria. *Annu. Rev. Cell Biol*. **6**: 297–327.

Lambowitz, A.M. and Belfort, M. (1993) Introns as mobile genetic elements. *Annu. Rev. Biochem*. **62**:

587–622.
Mizuuchi, K. (1992) Transpositional recombination: Mechanistic insights from bacteriophage Mu and other elements. *Annu. Rev. Biochem.* **61**: 1011–1051.
Sundaresan, V. (1996) Horizontal spread of transposon mutagenesis — new uses for old elements. *Trends Plant Sci.* **6**: 184–190.
Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**: 335–340.

第 14 章　诱变和 DNA 修复
Mutagenesis and DNA Repair

参考文献

Friedberg, E.C., Walker, G.C. and Siede, W. (1995) *DNA Repair and Mutagenesis*. ASM Press, Washington, DC.

推荐读物

Chu, G. and Mayne, L. (1996) Xeroderma-pigmentosum, Cockayne syndrome and trichothiodystrophy — do the genes explain the diseases? *Trends Genet.* **12**: 187–192.
Friedberg, E.C. (1996) Relationships between DNA repair and transcription. *Annu. Rev. Biochem.* **65**: 15–42.
Lindahl, T., Karran, P. and Wood, R.D. (1997) DNA excision repair pathways. *Curr. Opin. Genet. Dev.* **7**: 158–169.
Modrich, P. and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination and cancer biology. *Annu. Rev. Biochem.* **65**: 101–133.
Sancar, A. (1996) DNA excision repair. *Annu. Rev. Biochem.* **65**: 43–81.
Walker, G.C. (1995) SOS-regulated proteins in translesion DNA synthesis and mutagenesis. *Trends Biochem. Sci.* **20**: 416–420.
Wood, R.D. (1996) DNA repair in eukaryotes. *Annu. Rev. Biochem.* **65**: 135–167.

第 15 章　突变与选择
Mutation and Selection

参考文献

Cooper, D.N. and Krawczak, M. (1993) *Human Gene Mutation*. BIOS Scientific Publishers, Oxford.
Humphries, S. and Malcolm, S. (1994) *From Genotype to Phenotype*. BIOS Scientific Publishers, Oxford.
Li, W.-H. and Grauer, D. (1991) *Fundamentals of Molecular Evolution*. Sinauer, Sunderland, MA.

推荐读物

Britten, R.J. (1986) Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231**: 1393–1398.
Cao, A., Galanello, R. and Rosatelli, M.C. (1994) Genotype–phenotype correlations in β-thalassemias. *Blood Rev.* **8**: 1–12.
Drake, J.W. (1991) Spontaneous mutation. *Annu. Rev. Genet.* **25**: 125–146.
Miller, J.H. (1983) Mutational specificity in bacteria. *Annu. Rev. Genet.* **17**: 215–238.
Patel, P.I. and Lupski, J.R. (1994) Charcot–Marie–Tooth disease — a new paradigm for the mechanism of inherited disease. *Trends Genet.* **10**: 128–133.
Richards, R.I. and Sutherland, G.R. (1997) Dynamic mutation: Possible mechanisms and significance in human disease. *Trends Biochem. Sci.* **22**: 432–436.
Sniegowski, P.D. and Lenski, R.E. (1995) Mutation and adaption — the directed mutation controversy in evolutionary perspective. *Annu. Rev. Ecol. Systematics* **26**: 553–578.
Spencer, D.M. (1996) Creating conditional mutations in mammals. *Trends Genet.* **12**: 181–187.
Wagner, S.D. and Neuberger, M.S. (1996) Somatic hypermutation of immunoglobulin genes. *Annu. Rev. Immunol.* **14**: 441–457.

第 16 章　核酸结构
Nucleic Acid Structure

参考文献

Blackburn, G.M. and Gait, M.J. (eds) (1996) *Nucleic Acids in Chemistry and Biology*. Oxford University

Press, Oxford.
Dickerson, R.E. *et al.* (1989) *Definitions and nomencla*ture of nucleic acid structure parameters. *EMBO J.* **8**: 1–4.
Kornberg, A. and Baker, T.A. (1992) DNA structure and function. In: *DNA Replication*. 2nd edn, pp. 1–52. W.H. Freeman, New York.
Rich, A. *et al.* (1984) The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.* **53**: 791–846.
Yang, Y., Westcott, T.P., Pedersen, S.C., Tobias, I. and Olson, W.K. (1995) Effects of localised bending on DNA supercoiling. *Trends Biochem. Sci.* **20**: 313–319.

推荐读物

Doudna, J.A. and Cate, J.M. (1997) RNA structure: crystal clear? *Curr. Op. Struct. Biol.* **7**: 310–316.
Eaton, B.E. and Pieken, W.A. (1995) Ribonucleosides and RNA. *Annu. Rev. Biochem.* **64**: 837–863.
Frank-Kamenetskii, M.D. and Mirkin, S.M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.* **65**: 65–95.
Jaeger, J.A., Santa-Lucia, J. and Tinoco, I. (1993) Determination of RNA structure and thermodynamics. *Annu. Rev. Biochem.* **62**: 255–287.
Lebrun, A. and Lavern R. (1997) Unusual DNA conformations. *Curr. Op. Struct. Biol.* **7**: 348–354.
Roca, J. (1995) The mechanisms of DNA topoisomerases. *Trends Biochem. Sci.* **20**: 156–160.
Scott, W.G. and Klug, A. (1996) Ribozymes — structure and mechanism in RNA catalysis. *Trends Biochem. Sci.* **21**: 220–224.
Stark, W.M. and Boocock, M.R. (1995) Topological selectivity in site-specific recombination. In: *Mobile Genetic Elements: Frontiers in Molecular Biology* (ed. D. Sherratt), pp. 101–129. Oxford University Press, Oxford.
Strobel, S.A. and Doudna, J.A. (1997) RNA seeing double: Close-packing of helices in RNA tertiary structure. *Trends Biochem. Sci.* **22**: 262–266.
Travers, A.A. (1995) DNA bending by sequence and proteins. In: *DNA Protein: Structural Interactions: Frontiers in Molecular Biology* (Lilley, D.M.J. ed). Oxford University Press, Oxford, pp. 49–75.
Weeks, K.M. (1997) Protein-facilitated RNA folding. *Curr. Op. Struct. Biol.* **7**: 336–342.

第 17 章　核酸结合蛋白
Nucleic Acid-Binding Proteins

参考文献

Latchman, D.S. (1995) *Eukaryotic Transcription Factors. 2nd Edn.* Academic Press, London.
Liley, D. (ed.) (1995) *DNA–Protein: Structural Interactions: Frontiers in Molecular Biology.* Oxford University Press, Oxford.
Nagai, K. and Mattaj, I. (eds) (1995) *RNA–Protein Interactions: Frontiers in Molecular Biology.* Oxford University Press, Oxford.
Ptashne, M. (1992) *A Genetic Switch: Phage λ and Higher Organisms. 2nd Edn.* Cell Press, MA/Blackwell Science, MA.

推荐读物

Burley, S.K., Xie, X., Clark, K.L. and Shu, F. (1997) Histone-like transcription factors in eukaryotes. *Curr. Opin. Struct. Biol.* **7**: 94–102.
Choo, Y. and Klug, A. (1997) Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.* **7**: 117–125.
Chytil, M. and Verdine, G.L. (1996) The Rel family of eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* **6**: 91–100.
Harrison, S.C. (1991) A structural taxonomy of DNA-binding domains. *Nature* **353**: 715–719.
Nagashi, K. (1996) RNA–protein complexes. *Curr. Opin. Struct. Biol.* **6**: 53–61.
Pabo, C. and Saur, R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**: 1053–1095.
Raumann, B.E., Brown, B.M. and Sauer, R.T. (1994) Major groove DNA recognition by β-sheets: the ribbon-helix-helix family of gene regulatory proteins. *Curr. Opin. Struct. Biol.* **4**: 36–43.
Rhodes, D., Schwabe, J.W.R., Chapman, L and Fairall, L. (1996) Towards an understanding of protein-DNA recognition. *Phil. Trans. R. Soc.* **351**: 501–509.
Schwabe, J.W.R. (1997) The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* **7**: 126–134.
Siomi, H. and Dreyfuss, G. (1997) RNA-binding proteins as regulators of gene expression. *Curr. Opinion. Genet. Dev.* **7**: 345–353.
Suzuki, M. (1993). Common features in DNA recognition helices of eukaryotic transcription factors. *EMBO J.* **12**: 3221–3226.
Suzuki, M. and Giraldo, R. (1995) Zipperless bZips and zipped homeodomains. *Proc. Jpn Acad. Sci.* **71**: 39–44.

第18章 癌基因与癌
Oncogenes and Cancer

推荐读物

Bishop, J.M. (1995) Cancer — the rise of the genetic paradigm. *Genes Dev.* **9**: 1309–1315.

Herschman, H.R. (1991) Primary response genes induced by growth factors and tumour promoters. *Annu. Rev. Biochem.* **60**: 281–319.

Hunter, T. and Karin, M. (1992) The regulation of transcription by phosphorylation. *Cell* **70**: 375–387.

Jacks, T. (1996). Tumor suppressor gene mutations in mice. *Annu. Rev. Genet.* **30**: 603–636.

Lees, E.M. and Harlow, E. (1995) Cancer and the cell cycle. In: *Cell Cycle Control: Frontiers in Molecular Biology*, (eds C. Hutchinson and D.M. Glover), pp. 228–263. Oxford University Press, Oxford.

Lowy, D.R. (1993) Function and regulation of Ras. *Annu. Rev. Biochem.* **62**: 851–891.

Milner, J. (1995) Flexibility — the key to p53 function. *Trends Biochem. Sci.* **20**: 49–51.

Sherr, C.J. and Roberts, J.M. (1995) Inhibitors of mammalian G1 cyclin-dependent kinases. *Genes Dev.* **9**: 1149–1163.

Vaux, D.L. and Strasser, A. (1996) The molecular biology of apoptosis. *Proc. Natl. Acad. Sci. USA* **93**: 2239–2244.

Weinberg, R.A. (1996) E2F and cell proliferation — a world turned upside-down. *Cell* **85**: 457–459.

Weinberg, R.A. (1997) The cat and mouse games that genes, viruses, and cells play. *Cell* **88**: 573–575.

Zhang, C-C. (1996) Bacterial signalling involving eukaryotic-type protein kinases. *Mol. Microbiol.* **20**: 9–15.

第19章 细胞器基因组
Organelle Genomes

参考文献

Clayton, D.A. (1992) *Int. Rev. Cytol.* **141**: 217–232.

推荐读物

Larrson, N.G. and Clayton, D.A. (1995) Molecular-genetic aspects of human mitochondrial disorders. *Annu. Rev. Genet.* **29**: 151–178.

Mattson, M.P. (1997) Mother's legacy: Mitochondrial DNA mutations and Alzheimer's disease. *Trends Neurosci.* **20**: 373–375.

Shadel, G.S. and Clayton, D.A. (1997) Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* **66**: 409–435.

Stern, D.B., Higgs, D.C. and Yang, J.J. (1997) Transcription and translation in chloroplasts. *Trends Plant Sci.* **2**: 308–315.

第20章 质粒
Plasmids

推荐读物

Eguchi, Y., Itoh, T. and Tomizawa, J. (1991) Antisense RNA. *Annu. Rev. Biochem.* **60**: 631–652.

Hiraga, S. (1992) Chromosome and plasmid partition in *E. coli*. *Annu. Rev. Biochem.* **61**: 283–306.

Kornberg, A. and Baker, T.A. (1992) Plasmids and organelles. In: *DNA Replication (2nd edn)*, pp. 637–689. W.H. Freeman, New York.

Nordstron, K. and Austin, S.J. (1989) Mechanisms that contribute to the stable segregation of plasmids. *Annu. Rev. Genet.* **23**: 37–69.

Summers, D.K. (1996) *Plasmid Biology*. Blackwell Science, Oxford.

第21章 聚合酶链式反应(PCR)
The Polymerase Chain Reaction (PCR)

参考文献

Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds) (1990) PCR Protocols. *A Guide to Methods and*

Applications. Academic Press, San Diego, CA.

McPherson, M.P. and Hames, B.D. (eds) (1995) *PCR 2: A Practical Approach*. IRL Press, Oxford.

McPherson, M.P., Quirke, P. and Taylor, G.R. (eds) (1991) *PCR: A Practical Approach*. IRL Press, Oxford.

推荐读物

Audic, S. and Beraud-Colomb, E. (1997) Ancient DNA is 13 years old. *Nature Biotech*. 9: 855–858.

Caetanoanolles, G. (1996) Scanning nucleic acids by *in vitro* amplification — new developments and applications. *Nature Biotech*. 14: 1668–1674.

Erlich, H.A. and Arnheim, N. (1992) Genetic analysis using the polymerase chain reaction. *Annu. Rev. Genet*. 26: 479–506.

Jansson, J.K. and Prosser, J.I. (1997) Quantification of the presence and activity of specific microorganisms in nature. *Mol. Biotech*. 7: 103–120.

McClelland, M., Mathieudauda, F. and Welsh, J. (1995) RNA fingerprinting and differential display using arbitrarily primed PCR. *Trends Genet*. 11: 242–246.

Nanda, S.K. and Jain, S.K. (1995) *In vitro* nucleic acid amplification systems. *Curr. Sci*. 66: 421–429.

Ohan, N.W. and Heikkila, J.J. (1995) Reverse transcription polymerase chain reaction — an overview of the technique and its applications. *Biotech. Adv*. 11: 13–29.

Pena, S.D.J. and Chakraborty, R. (1994) Paternity testing in the DNA era. *Trends Genet*. 10: 352–356.

Powell, W., Machray, G.C. and Provan, J. (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci*. 7: 215–222.

Register, J.C. (1997) Approaches to evaluating the transgenic status of transformed plants. *Trends Biotech*. 15: 141–146.

Reischl, U. and Kochanowski, B. (1995) Quantitative PCR — a survey of the present technology. *Mol. Biotech*. 3: 55–71.

Sagerstrom, C.G., Sun, B.I. and Sive, H.L. (1997) Subtractive cloning: Past, present and future..*Annu. Rev. Biochem*. 66: 751–783.

Slatko, B.E. (1996) Thermal cycle dideoxy sequencing. *Mol. Biotech*. 6: 311–322.

Whelen, A.C. and Persing, D.H. (1996) The role of nucleic acid amplification and detection in the clinical microbiology laboratory. *Annu. Rev. Microbiol*. 50: 349–373.

第 22 章　蛋白质:结构,功能与进化
Proteins: Structure, Function and Evolution

参考文献

Branden, C. and Tooze, J. (1991) *Introduction to Protein Structure*. Garland Press, New York.

Doolittle, R.F. (1985) Proteins. *Sci. Amer*. 253: 88–99.

Lewin, B. (1996) *Genes VI*. Oxford University Press, Oxford, pp. 663–711.

Richardson, J.S., Richardson, D.C., Tweedy, N.B., Gernet, K.M., Quinn, T.P., Hecht, M.H., Erickson, B.W., Yan, Y., McClain, R.D., Donlan, M.E. and Suries, M.C. (1992) Looking at proteins: representations, folding, packing and design. *Biophys. J*. 63: 1186–1220.

Stryer, L. (1995) *Biochemistry* (*4th edn*). W.H. Freeman, New York.

推荐读物

Allen, J.B., Wallberg, M.W., Edwards, M.C. and Elledge, S.J. (1995) Finding prospective partners in the library — the 2-hybrid system and phage display find a match. *Trends Biochem. Sci*. 20: 511–516.

Britten, R.J. (1996) DNA sequence insertion and evolutionary variation in gene regulation. *Proc. Natl Acad. Sci. USA* 93: 9374–9377.

Cooke, J., Nowak, M.A., Boerlijst, M., and Maynard-Smith, J. (1997) Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends Genet*. 13: 360–363.

Cooper, A.A. and Stevens, T.H. (1995) Protein splicing — self-splicing of genetically mobile elements at the protein level. *Trends Biochem. Sci*. 20: 351–356.

Davidson, J.N. and Peterson, M.L. (1997) Origin of genes encoding multi-enzyme proteins in eukaryotes. *Trends Genet*. 13: 281–285.

Dibb, N.J. (1993) Why do genes have introns? *FEBS Lett*. 325: 135–139.

Doolittle, R.F. (1995) The origins abd evolution of eukaryotic proteins. *Phil. Trans. R. Soc. Lond*. 349: 235–240.

Evans, M.J., Carlton, M.B.L. and Russ, A.P. (1997) Gene trapping and functional genomics. *Trends Genet*. 13: 370–375.

Fersht, A.R. (1997) Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol*. 7: 3–9.

Gabor Miklos, G.L. and Rubin, G.M. (1996) The role of the genome projects in determining gene function: Insights from model organisms. *Cell* 86: 521–529.

Goeffeau, A., Barrell, B.G., Bussey, *et al.* (1996) Life with 6000 genes. *Science* **274**: 546–567.
Gorlich, D. and Mattaj, I.W. (1996) Nucleocytoplasmic transport. *Science* **271**: 1513–1518.
Govindarajan, S. and Goldstein, R.A. (1996) Why are some protein structures so common? *Proc. Natl Acad. Sci. USA* **98**: 3341–3345.
Heinkoff, S., Greene, E.A., Pietrokovski, S., Bork, P., Attwood, T.K. and Hood, L. (1997) Gene families: the taxonomy of protein paralogs and chimeras. *Science* **278**: 609–614.
Hieter, P. and Boguski, M. (1997) Functional genomics: it's all how you read it. *Science* **278**: 601–602.
Hochstrasser, M. (1996) Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**: 405–439.
Holland, P.W.H., Garcia-Fernandez, J., Williams, N.A. and Sidow, A. (1994) Gene duplications and the origins of vertebrate development. *Development (Suppl.)*: 125–133.
Lander, E.S. (1996) The new genomics: Global views of biology. *Science* **274**: 536–539.
Levitt, M., Gerstein, M., Huang, E., Subbiah, S. and Ysai, J. (1997) Protein folding — the endgame. *Annu. Rev. Biochem.* **66**: 549–579.
Loomis, W.F. and Sternberg, P.W. (1995) Genetic Networks. *Science* **269**: 649.
Martin, J. and Hartl, F.U. (1997) Chaperone-assisted protein folding. *Curr. Opin. Struct. Biol.* **7**: 41–52.
Neupert, W. (1997) Protein import into mitochondria. *Annu. Rev. Biochem.* **66**: 863–917.
Rapoport, T.A., Jungnickel, B. and Kutay, U. (1996) Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65**: 801–848.
Rechsteiner, M. and Rogers, S.W. (1996) PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21**: 267–271.
Roder, H. and Colon, W. (1997) Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**: 15–28.
Rothman, J.E. (1994) Mechanisms of intracellular transport. *Nature* **372**: 55–68.
Schatz, G. and Dobberstein, B. (1996) Common principles of protein translocation across membranes. *Science* **271**: 1519–1526.
Tatusov, R.L., Koonin, E.V. and Lipman, D.J. (1997) A genomic perspective on protein families. *Science* **278**: 631–637.
Walter, P. and Johnson, A.E. (1994) Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10**: 87–119.
Xue, L. and Noll, M. (1996) The functional conservation of proteins in evolutionary alleles and the dominant role of enhancers in evolution. *EMBO J.* **15**: 3722–3731.

第23章　蛋白质合成
Protein Synthesis

推荐读物

Farabaugh, P.J. (1996) Programmed translational frameshifting. *Annu. Rev. Genet.* **30**: 507–528.
Green, R. and Noller, H.F. (1997) Ribosomes and translation. *Annu. Rev. Biochem.* **66**: 679–716.
Jacobson, A. and Peltz, S.W. (1996) Interrelationships of the pathways of messenger-RNA decay and translation in eukaryotic cells. *Annu. Rev. Biochem.* **65**: 693–739.
Mountford, P.S. and Smith, A.G. (1995) Internal ribosome entry sites and dicistronic RNAs in mammalian transgenesis. *Trends Genet.* **11**: 179–184.
Ruizechevarria, M.J., Czaplinski, K. and Peltz, S.W. (1996) Making sense of nonsense in yeast. *Trends Biochem. Sci.* **21**: 433–438.

第24章　重组DNA与分子克隆
Recombinant DNA and Molecular Cloning

参考文献

Brown, T.A. (ed.) *DNA Cloning: a Practical Approach. 2nd Edn.* (*4 volumes*). IRL Press, Oxford
Brown, T.A. (ed.) *Essential Molecular Biology: a Practical Approach* (*2 volumes*). IRL Press, Oxford.
Old, R.W. and Primrose, S.B. (1996) *Principles of Gene Manipulation: an Introduction to Genetic Engineering.* Blackwell Science, Oxford.
Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (*3 volumes*). Cold Spring Harbor Press, Cold Spring Harbor, NY.
Watson, J.D., Gilman, M., Witkowski, J. and Zoller, M. (1992) *Recombinant DNA. 2nd Edn.* Scientific American Books, New York.

推荐读物

Allen, J.B., Wallberg, M.W., Edwards, M.C. and Elledge, S.J. (1995) Finding prospective partners in

the library — the 2-hybrid system and phage display find a match. *Trends Biochem. Sci.* 20: 511–516.

Anderson, W.F. (1992) Human gene therapy. *Science* 256: 808–813.

Beddington, R. (1992) Transgenic mutagenesis in the mouse. *Trends Genet.* 8: 345–347.

Coutre, L.A. and Stinchcomb, D.T. (1996) Anti-gene therapy: the use of ribozymes to inhibit gene function. *Trends Genet.* 12: 510–515.

Evans, M.J., Carlton, M.B.L. and Russ, A.P. (1997) Gene trapping and functional genomics. *Trends Genet.* 13: 370–375.

Jain, S.M. (1993) Recent advances in plant genetic engineering. *Curr. Sci.* 64: 715–724.

Monaco, A.P. and Larin, Z. (1994) YACs, BACs, PACs and MACs — Artificial chromosomes as research tools. *Trends Biotech.* 12: 280–286.

Mountford, P.S. and Smith, A.G. (1995) Internal ribosome enty sites and dicistronic RNAs in mammalian transgenesis. *Trends Genet.* 11: 179–184.

Nielsen, P.E. (1996) Peptide nucleic acids — a new dimension to peptide libraries and aptames. *Methods Enzymol.* 267: 426–433.

Parimoo, S., Patanjali, S.R., Kolluri, R., Xu, H.X., Wei,

第25章　重组
Recombination

推荐读物

Cameriniotero, R.D. and Hsieh, P. (1995) Homologous recombination proteins in prokaryotes and eukaryotes. *Annu. Rev. Genet.* 29: 509–552.

Eggleston, A.K. and West, S.C. (1996) Exchanging partners — recombination in *Escherichia coli*. *Trends Genet.* 12: 20–26.

Hagmann, M. (1997) RAGged repair: What's new in V(D)J recombination. *Biol. Chem.* 378: 815–819.

Lichten, M. and Goldman, A.S.H. (1995) Meiotic recombination hotspots. *Annu. Rev. Genet.* 29: 423–444.

Rao, B.J., Chiu, S.K., Bazemore, L.R., Reddy, G. and Radding, C.M. (1995) How specific is the first recognition step of homologous recombination? *Trends Biochem. Sci.* 20: 109–113.

Shinagawa, H. and Iwasaki, H. (1996) Processing the Holliday junction in homologous recombination. *Trends Biochem. Sci.* 21: 107–111.

Stavnezer, J. (1996) Antibody class switching. *Adv. Immunol.* 61: 79–146.

第26章　复制
Replication

参考文献

Baker T.A. and Wickner S.H. (1992) Genetics and enzymology of DNA replication in *E. coli*. *Annu. Rev. Genet.* 26: 447–477.

De Pamphilis M.L. (Ed.) (1996) *DNA Replication in Eukaryotic Cells*. Cold Spring Harbor Press, New York.

Kornberg A. and Baker T.A. (1992) *DNA Replication*. W.H. Freeman, New York.

推荐读物

Chong, J.P.J., Thommes, P. and Blow, J.J. (1996) The role of MCM/P1 proteins in the licensing of DNA replication. *Trends Biochem. Sci.* 21: 102–106.

Sousa, R. (1996) Structural and mechanistic relationships between nucleic acid polymerases. *Trends Biochem. Sci.* 21: 186–190.

Wold, M.S. (1997) Replication protein A: A heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* 66: 61–92.

Wynford-Thomas, D. and Kipling, D. (1997) Telomerase: Cancer and the knockout mouse. *Nature* 389: 551–552.

第27章　RNA 加工
RNA Processing

推荐读物

Bachellerie, J.P. and Cavaille, J. (1997) Guiding ribose methylation of rRNA. *Trends Biochem. Sci.* 22: 257–261.

Bass, B.L. (1997) RNA editing and hypermutation by

adenosine deamination. *Trends Biochem. Sci.* **22**: 157–162.

Chabot, B. (1996) Directing alternative splicing — cast and scenarios. *Trends Genet.* **12**: 472–478.

Chen, C.Y.A. and Shyu, A.B. (1995) AU-rich elements — characterization and importance in messenger-RNA degradation. *Trends Biochem. Sci.* **20**: 465–470.

Copertino, D.W. and Hallick, R.B. (1995) Group II and group III introns of twintrons — potential relationships with nuclear premessenger RNA introns. *Trends Biochem. Sci.* **18**: 467–471.

Guo, Z.J. and Sherman, F. (1996) 3'-end-forming signals of yeast mRNA. *Trends Biochem. Sci.* **21**: 477–481.

Kable, M.L., Heidmann, S. and Stuart, K.D. (1997). RNA editing: Getting U into RNA. *Trends Biochem. Sci.* **22**: 162–166.

Ross, J. (1996) Control of messenger-RNA stability in higher eukaryotes. *Trends Genet.* **12**: 171–175.

Sarker, N. (1997) Polyadenylation of mRNA in prokaryotes. *Annu. Rev. Biochem.* **66**: 173–197.

Tarn, W.Y. and Steitz, J.A. (1997) Pre-mRNA splicing: The discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* **22**: 132–137.

Valcarcel, J. and Green, M.R. (1996) The SR protein family — pleiotrophic functions in pre-messenger-RNA splicing. *Trends Biochem. Sci.* **21**: 296–301.

Wahle, E. and Keller, W. (1996) The biochemistry of polyadenylation. *Trends Biochem. Sci.* **21**: 247–250.

第 28 章 信号转导
Signal Transduction

推荐读物

Barford, D. (1996) Molecular mechanisms of the protein serine threonine phosphatases. *Trends Biochem. Sci.* **21**: 407–412.

Bourne, H.R. (1997) How receptors talk to trimeric G proteins. *Curr. Opin. Cell Biol.* **9**: 134–142.

Cohen, C.B., Ren, R. and Baltimore, D. (1995) Molecular binding domains in signal transduction proteins. *Cell* **80**: 237–248.

Cohen, P.T.W. (1997) Novel protein serine/threonine phosphatases: Variety is the spice of life. *Trends Biochem. Sci.* **22**: 245–251.

Fauman, E.B. and Saper, M.A. (1996) Structure and function of the protein tyrosine phosphatases. *Trends Biochem. Sci.* **21**: 413–417.

Houslay, M.D. and Milligan, G. (1997) Tailoring cAMP-signalling responses through isoform muliplicity. *Trends Biochem. Sci.* **22**: 217–224.

Jan, L.Y. and Jan, Y.N. (1997) Receptor-regulated ion channels. *Curr. Opin. Cell Biol.* **9**: 155–160.

Lohmann, S.M., Vaandrager, A.B., Smolenski, A., Walter, U. and de Jonge, H.R. (1997) Distinct and specific functions of cAMP-dependent protein kinases. *Trends Biochem. Sci.* **22**: 307–312.

Marshall, C.J. (1995) Specificity of receptor tyrosine kinase signalling: Transient verses sustained extracellular signal-regulated kinase activation. *Cell* **80**: 179–186.

Morrison, D.K. and Cutler, R.E. Jr. (1997) The complexity of Raf-1 regulation. *Curr. Opin. Cell Biol.* **9**: 174–179.

Newton, A.C. (1997) Regulation of protein kinase C. *Curr. Opin. Cell Biol.* **9**: 161–167.

Pellegrini, S. and Dusanter-Fourt, I. (1997) The structure, regulation and function of the Janus kinases (JAKs) and the signal transducers and activators of transcription (STATs). *Eur. J. Biochem.* **248**: 615–633.

Robinson, M.J. and Cobb, M.H. (1997) Mitogen-activated kinase pathways. *Curr. Opin. Cell Biol.* **9**: 180–186.

Singer, W.D., Brown, H.A. and Sternweis, P.C. (1997) Regulation of eukaryotic phosphatidylinositol-specific phospholipase C and phosphoslipase D. *Annu. Rev. Biochem.* **66**: 475–509.

Spiegel, S., Foster, D. and Kolesnick, R. (1996) Signal transduction through lipid second messengers. *Curr. Opin. Cell Biol.* **8**: 159–167.

Sprang, S.R. (1997) G protein mechanisms: Insights from structural analysis. *Annu. Rev. Biochem.* **66**: 639–678.

Ten Dijke, P., Miyazono, K. and Heldin, C.-H. (1996) Signalling via hetero-oligomeric complexes of type I and type II serine/threonine kinase receptors. *Curr. Opin. Cell Biol.* **8**: 139–145.

Toker, A. and Cantley, L.C. (1997) Signalling through the lipid products of phosphoinositide-3-OH kinase. *Nature* **387**: 673–676.

Van Haesebroeck, B., Leevers, S.J., Panayotou, G. and Waterfield, M.D. (1997) Phosphoinositide 3-kinases: A conserved family of signal transducers. *Trends Biochem. Sci.* **22**: 267–272.

Whitman, M. (1997) Feedback from inhibitory SMADs. *Nature* **398**: 549–551.

Wittinghofer, A. and Nassar, N. (1997) How Ras-related proteins talk to their effectors. *Trends Biochem. Sci.* **21**: 488–491.

第 29 章 转录
Transcription

推荐读物

Adhya, S. (1996) The *lac* and *gal* operons today. In: *Regulation of Gene Expression in Escherichia coli* (eds

E.C.C. Lin and A.S. Lynch), pp. 181–200. R.G. Landes, Texas/Chapman & Hall, New York.
Bjorklund, S. and Kim, Y.J. (1996) Mediator of transcriptional regulation. *Trends Biochem. Sci.* **21**: 335–337.
Busby, S. and Kolb, A. (1996) The Cap modulon. In: *Regulation of Gene Expression* in *Escherichia coli* (eds E.C.C. Lin and A.S. Lynch), pp. 255–280. R.G Landes, Texas/Chapman & Hall, New York.
Hannarose, W. and Hansen, U. (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet.* **12**: 229–234.
Henkin, T.M. (1996) Control of transcriptional termination in prokaryotes. *Annu. Rev. Genet.* **30**: 35–57.
Kamakaka, R.T. (1997) Silencers and locus control regions: Opposite sides of the same coin. *Trends Biochem. Sci.* **22**: 124–128.
Magasanik, B. (1996) Regulation of nitrogen assimilation. In: *Regulation of Gene Expression* in *Escherichia coli* (eds E.C.C. Lin and A.S. Lynch), pp. 281–290 R.G. Landes, Texas/Chapman & Hall, New York .
Novina, C.D. and Roy, A.L. (1996) Core promoters and transcriptional control. *Trends Genet.* **12**: 351–355.
Ptashne, M. and Gann, A. (1997) Transcriptional activation by recruitment. *Nature* **386**: 569–575.
Reeder, R.M. and Lang, W.H. (1997) Terminating transcription in eukaryotes: lessons from RNA polymerase I *Trends Biochem. Sci.* **22**: 473–477.
Reines, D.,.Conaway, J.W. and Conaway, R.C. (1996) The RNA polymerase II general elongation factors. *Trends Biochem. Sci.* **21**: 351–355.
Rippe, K., Von Hippel, P.H. and Langowski, J. (1995) Action at a distance — DNA-looping and initiation of transcription. *Trends Biochem. Sci.* **20**: 500–506.
Roberts, J.W. (1996) Transcription termination and its control. In: *Regulation of Gene Expression* in *Escherichia coli* (eds E.C.C. Lin and A.S. Lynch), pp. 27–46. R.G. Landes, Texas/Chapman & Hall, New York.
Roeder, R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**: 327–335.
Svaren, J. and Horz, W. (1997) Transcription factors *vs* nucleosomes: Regulation of the *PHO5* promoter in yeast. *Trends Biochem. Sci.* **22**: 93–97.
Svejstrup, J.Q., Vichi, P. and Egly, J.M. (1996) The multiple roles of transcription/repair factor TFIIH. *Trends Biochem. Sci.* **21**: 346–350.
Uptain, S.M., Kane, C.M. and Chamberlin, M.J. (1997) Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* **66**: 117–172.
Van Dromme, M., Gauthier-Rouviere, C., Lamb, N. and Fernandez, A. (1996) Regulation of transcription factor localisation — fine-tuning of gene expression. *Trends Biochem. Sci.* **21**: 59–64.
Verrijzer, C.P. and Tjian, R. (1996) TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.* **21**: 338–342.
Wood, W.G. (1996) The complexities of β-globin gene regulation. *Trends Genet.* **12**: 204–206.

第 30 章 病毒与亚病毒感染因子
Viruses and Subviral Agents

参考文献

Dimmock, N.J. and Primrose, S.B. (1993) *An Introduction to Modern Virology.* 4th edn. Blackwell Science, Oxford.
Murphy, F.A., Fauquet, C.M., Bishop D.H.L, Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A. and Summers, M.D. (eds) (1995) *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses.* Springer, New York.

推荐读物

Aguzzi, A. and Weissmann, C. (1997) Prion research: the next frontiers. *Nature* **389**: 795–798.
Banerjee, A.K. and Barik, S. (1992) Gene expression of vesicular stomatosis virus genome. *Virology* **188**: 417–429.
Berns, K.I. (1990) Parvovirus replication. *Microbiol. Rev.* **54**: 316–329.
Caughey, B. and Chesebro, B. (1997) Prion protein and the transmissible spongiform encephalopathies. *Trends Cell Biol.* **7**: 56–62.
Coffin, J.M. (1992) Structure and classification of retroviruses. In: *The Retroviridae* (ed. J. Levy), vol. 1, pp. 1437–1500. Plenum Press, New York.
Collmer, C. and Howell, S. (1992) Role of satellite RNA in the expression of symptoms caused by plant viruses. *Annu. Rev. Phytopathol.* **30**: 419–442.
Diener, T.O. (1991) The frontiers of life: The viroids and viroid-like satellite RNAs. In: *Viroids and Satellites: Molecular Parasites at the Frontiers of Life* (ed. K. Maramorosch), pp. 1–20. CRC Press, Boca Raton, FL.
Eckhart, W. (1991) Polyomavirinae and their replication. In: *Fundamental Virology* (eds B.N. Fields and D.M. Knipe), 2nd edn., pp. 727–741. Raven Press, New York.
Estes, M.K. (1991) Rotaviruses and their replication. In: *Fundamental Virology* (eds B.N. Fields and D.M. Knipe), 2nd edn., pp. 619–642. Raven Press, New York.
Gabizon, R. and Taraboulos, A. (1997) Of mice and (mad) cows: Transgenic mice help to understand prions. *Trends Genet.* **13**: 264–269.
Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (eds) (1983) *Lambda II.* Cold Spring Harbor Press, Cold Spring Harbor, New York.
Horwitz, M.S. (1990) Adenoviruses and their replication. In: *Virology* (eds B.N. Fields and D.M. Knipe), 2nd edn., pp. 1679–1722. Raven Press, New York.
Joshi, S. and Joshi, R.L. (1996) Molecular biology of Human Immunodeficiency Virus type 1. *Transfusion Sci.* 17: 351–378.

Keppel, F., Fayewt, O. and Georgopoulos K. (1988) Strategies for bacteriophage DNA replication. In: *The Bacteriophages* (ed. R. Calendar), Vol. 2, pp. 145–262. Plenum Press, New York.

Kingsbury, D. (ed.) (1991) *The Paramyxoviruses*. Plenum Press, New York.

Luong, G. and Palese, P. (1992) Genetic analysis of influenza virus. *Curr. Opin. Genet. Dev.* 2: 77–81.

Roizman, B. and Sears, A.E. (1993) The replication of Herpes simplex viruses In: *The Human Herpesvirus* (eds B. Roizman, C. Lopez and R.J. Whitley), pp. 11–68. Raven Press, New York.

第1章参考网站

On-line Mendelian Inheritance in Man (OMIM): http://gdbwww.gdb.org/omimdoc/omimtop.html

第12章参考网站

Genome mapping databases

Bacteria

H. influenzae — http://www.tigr.org/tdb/mdb/hidb/hidb.html

M. genitalium — http://www.tigr.org/tdb/mdb/mgdb/mgdb.html

E. coli — http://www.ncbi.nlm.nih.gov/Complete_Genomes/Ecoli/

B. subtilis — http://acnuc.univ-lyon1.fr/nrsub/nrsub.html

Yeast

S. cerevisiae — http://speedy.mips.biochem.mpg.de/mips/yeast/

Invertebrates

C. elegans — http://probe.nalusda.gov:8000/other/aboutacedb.html

D. melanogaster — http://www.embl-ebi.ac.uk/flybase/

Plants

A. thaliana — http://genome-www.stanford.edu/Arabidopsis/

Mammals

Humans (On-line Mendelian Inheritance in Man) — http://gdbwww.gdb.org/omimdoc/omimtop.html

Mouse — http://www.jax.org/

Nucleic acid and protein sequence databases

GenBank (DNA and protein sequences) — http://www.ncbi.nim.nih.gov

EMBL (DNA sequences) — http://www.ebi.ac.uk

DDBJ (DNA database of Japan) — http://www.nig.oc.jp/home.html

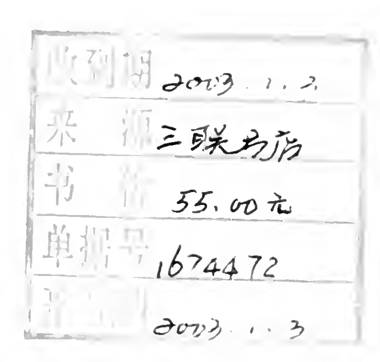
第30章参考网站

All the virology on the web: an index of virology resources and information available on the internet — http://www-micro.msb.le.ac.uk/garryfavweb/garryfavweb.html

索　　引

（中文按汉语拼音排序。t:表格,f:图,bx:文框）




收到期 2003.1.2
来源 三联书店
书价 55.00元
单据号 1674472
开票日期 2003.1.3


中科院植物所图书馆
S0000097

58.178
487

注　　意

1 借书到期请即送还。

2 请勿在书上批改圈点，折角。

3 借去图书如有污损遗失等情形须照章赔偿。

京卡0701

27479

(Q-0946.0103)
责任编辑：马学海　封面设计：王　浩

生命科学编辑部
联系电话：010-64012501
http://spbio.yeah.net
e-mail:spbio@163.net

A D V A N C E D
MOLECULAR BIOLOGY

高级分子生物学要义



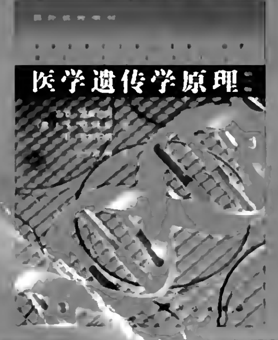
书　名　神经元：细胞和分子生物学（翻译版）　　2001年9月出版
Neuron: Cell and Molecular Biology　　估　价 55.00 元
作　者　〔美〕I.B.莱维坦，L.K.卡兹玛克　著
译　者　舒斯云　包新民　等
书　号　7-03-009098-9/Q·1036

本书是一本有着广泛影响的神经生物学名著。它从细胞及分子生物学的角度，介绍了脑的结构和活动规律，论述了神经细胞及神经胶质细胞的细胞生物学特征、细胞内和细胞间的信息传导、脑高级功能、神经元的生长发育与分化、可塑性等等。读者在掌握了本书的内容后，就能较容易地对神经系统的结构、发育、生理、行为等开展专门研究。
本书可作为神经生物学的研究生教材，也可作为从事神经科学研究的科学工作者的参考书。



书　名　微生物生物学（翻译版）　　2001年8月出版
作　者　〔美〕M. T. 马迪根　J. M. 马丁克　J. 帕克　　定　价 160.00元
译　者　杨文博　等
书　号　7-03-008345-8/Q·954

Brock Biology of Microorganisms 是国外的经典教科书。本书系其第八版的中文翻译版本，包括23章、附录及词汇表。该书从细胞化学和细胞生物学入手，详尽地介绍了微生物的营养、代谢、合成；分子遗传和微生物遗传；遗传工程和生物工程；生长和调控；工业微生物；寄主和寄生菌的关系；免疫学和免疫；临床和诊断微生物；流行病学和公共卫生微生物学；主要的微生物疾病；微生物代谢的多样性；微生物生态学；分子体系和微生物进化；病毒、细菌、古细菌和真核微生物等内容。该书阐述清晰，条理性强，简明易懂，并附有图表，每节后有概括的内容总结，便于读者学习。



书　名　医学遗传学原理（翻译版）　　2001年3月出版
Principles of Medical Genetics　　定　价 42.00元
作　者　〔美〕T. D. 盖莱哈特　F. S. 柯林斯　D. 金斯伯格
译　者　孙开来　等
书　号　7-03-008746-1/R·586

本书是世界著名的遗传学家，美国NIH人类基因组研究所主任F. S. Collins和密西根大学T. D. Gelehrter教授等编著的医学遗传学名著。近年来人类分子细胞遗传学迅猛发展，尤其在人类基因组计划方面取得了重大成就。作者运用这些新成就，结合遗传学基本理论，系统而有重点地阐述了遗传学在医学中的重要应用；书后还附有每章的问题与答案，有助于加深对内容的理解。译文较好地保持了原书的风格。本书是一本值得推荐的教科书和参考用书。适合作为医科院校的本科教材，也可供临床医生和医学遗传学研究人员及相关研究人员参考使用。



ISBN 7-03-008299-0
定　价：55.00 元

